PREPUBLICATION COPY

SCIENCE AND JUDGMENT IN RISK ASSESSMENT

ADVANCE COPY
Not For Public Release Before:
Wednesday, January 19, 1994
11:00 a.m. EST

NATIONAL RESEARCH COUNCIL

SCIENCE AND JUDGMENT IN RISK ASSESSMENT

COMMITTEE ON RISK ASSESSMENT
OF HAZARDOUS AIR POLLUTANTS

BOARD ON ENVIRONMENTAL STUDIES
AND TOXICOLOGY

COMMISSION ON LIFE SCIENCES

NATIONAL RESEARCH COUNCIL

PREPUBLICATION COPY

NATIONAL ACADEMY PRESS
WASHINGTON, D.C. 1994

NATIONAL ACADEMY PRESS 2101 Constitution Ave., N.W. Washington, D.C. 20418

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competencies and with regard for appropriate balance.

This report has been reviewed by a group other than the authors according to procedures approved by a Report Review Committee consisting of members of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine.

The National Academy of Sciences is a private, non-profit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Bruce M. Alberts is president of the National Academy of Sciences.

The National Academy of Engineering was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Robert M. White is president of the National Academy of Engineering.

The Institute of Medicine was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Kenneth I. Shine is president of the Institute of Medicine.

The National Research Council was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Bruce Alberts and Dr. Jack Halpern are chairman and vice chairman, respectively, of the National Research Council.

The project was supported by the U.S. Environmental Protection Agency under contract CR818293-01-0.

Additional copies of this report are available from the National Academy Press, 2101 Constitution Avenue, N.W., Washington, D.C. 20418

Printed in the United States of America

COMMITTEE ON RISK ASSESSMENT OF HAZARDOUS AIR POLLUTANTS

KURT J. ISSELBACHER *(Chairman)*, Massachusetts General Hospital, Charlestown, Mass.
ARTHUR C. UPTON *(Vice-Chairman)*, New York University Medical Center (retired), N.Y.
JOHN C. BAILAR, McGill University School of Medicine, Montreal, Canada
KENNETH B. BISCHOFF, University of Delaware, Newark, Del.
KENNETH T. BOGEN, Lawrence Livermore National Laboratory, Livermore, Calif.
JOHN I. BRAUMAN, Stanford University, Calif.
DAVID D. DONIGER, Natural Resources Defense Council, Washington, D.C. *
JOHN DOULL, University of Kansas Medical Center, Kansas City, Kan.
ADAM M. FINKEL, Resources for the Future, Washington, D.C.
CURTIS C. HARRIS, National Cancer Institute, Bethesda, Md.
PHILIP K. HOPKE, Clarkson University, Potsdam, N.Y.
SHEILA S. JASANOFF, Cornell University, Ithaca, N.Y.
ROGER O. MCCLELLAN, Chemical Industry Institute of Toxicology, Research Triangle Park, N.C.
LINCOLN E. MOSES, Stanford University, Calif.
D. WARNER NORTH, Decision Focus, Inc., Mountain View, Calif.
CRAIG N. OREN, Rutgers University School of Law, Camden, N.J.
REBECCA T. PARKIN, Beccam Services, Plainsboro, N.J.
EDO D. PELLIZZARI, Research Triangle Institute, N.Car.
JOSEPH V. RODRICKS, Environ Corporation, Arlington, Va.
ARMISTEAD G. RUSSELL, Carnegie-Mellon University, Pittsburgh, Penn.
JAMES N. SEIBER, University of Nevada, Reno, Nev.
STEVEN N. SPAW, Law Environmental Incorporated, Austin, Tex.
JOHN D. SPENGLER, Harvard University, Boston, Mass.
BAILUS WALKER, University of Oklahoma, Oklahoma City, Okla.
HANSPETER WITSCHI, University of California, Davis, Calif.

Staff

RICHARD D. THOMAS, Project Director
DEBORAH D. STINE, Study Director (until May 1993)
MARVIN A. SCHNEIDERMAN, Senior Scientist
NORMAN GROSSBLATT, Editor
ANNE M. SPRAGUE, Information Specialist (until August 1993)
RUTH E. CROSSGROVE, Information Specialist
RUTH P. DANOFF, Project Assistant (until October 1992)
SHELLEY A. NURSE, Senior Project Assistant (until February 1993)
CATHERINE M. KUBIK, Senior Project Assistant

*Left committee in May 1993 upon becoming Deputy Director of the White House Office of Environmental Quality

BOARD ON ENVIRONMENTAL STUDIES AND TOXICOLOGY

Staff

COMMISSION ON LIFE SCIENCES

PREFACE

Concerns over air and its contamination by toxic pollution have led to a body of legislation about clean air. The most recent of these pieces of legislation is the Clean Air Act Amendments of 1990. As part of that legislation, the Congress directed the administrator of the Environmental Protection Agency (EPA) to engage the National Academy of Sciences (NAS) in a review of the methods that EPA uses to ensure that Americans have appropriately clean air.

Congress's charge to the National Research Council (NRC) can be summarized in a short set of questions:

1. Given that quantitative risk assessment is essential for EPA's implementation of the Clean Air Act, is EPA conducting risk assessments in the best possible manner?
2. Has EPA developed mechanisms for keeping its risk-assessment procedures current in the face of new developments in science?
3. Are adequate data being collected to permit EPA to carry out its mandate?
4. What, if anything, should be done to improve EPA's development and use of risk assessments?

This report attempts to address these questions and to provide advice to enable EPA to carry out the mandates of the Clean Air Act and to satisfy Congress's concerns about the implementation of the act and its 1990 amendments. Specifically, the committee was charged with reviewing the risk-assessment methods used by EPA "to determine the carcinogenic risk associated with exposure to hazardous air pollutants" and to suggest improvements in such methods. The elements to be studied by the committee included "the techniques used for estimating and describing the carcinogenic potency to humans of hazardous air pollutants, as well as methods for estimating exposure to these materials." The legislation instructed NAS to "evaluate and report on the methodology for assessing the risk of adverse health effects" for hazardous air pollutants.

To meet the congressional mandate, and in direct response to the request from the administrator of EPA, NAS, through its research arm, the National Research Council, established the Committee on Risk Assessment of Hazardous Air Pollutants under the administrative umbrella of the Board on Environmental Studies and Toxicology. The committee consisted of 23 members and the disciplines represented included medicine, epidemiology, chemistry, chemical engineering, environmental health, law, pharmacology and toxicology, risk assessment, risk management, occupational health, statistics, air monitoring,

and public health. It included academics, industry scientists, public advocates, and state and local public-health officials.

The first meeting of the committee was held on October 31, 1991. In the course of the first several meetings, presentations were made to the committee by committee members and by individuals or representatives of groups with special concerns in the development and use of risk assessment. Among the latter were presenters on behalf of the American Industrial Health Council (an industry group), the Chemical Manufacturers Association, the American Petroleum Institute, the American Iron and Steel Institute, the American Chemical Society, such official public-health groups as the Texas Air Control Board and the State and Territorial Air Pollution Program Administrators, and such public-interest groups as the Natural Resources Defense Council and the Environmental Defense Fund. Presentations were also made by the legal representative of a paint manufacturer and by a senior member of an environmental consulting company.

Early in the course of its deliberations the committee developed a set of issues for consideration and reply by the EPA Office of Air and Radiation and Office of Research and Development. EPA's responses were presented to the committee both orally and in written form during the committee's meetings in late March 1992.

James Powell, of the U.S. Senate staff, provided the committee with both the legislative history of the Clean Air Act Amendments and the concerns of senators in the evolution of EPA's development of regulations. Greg Wetstone, of the U.S. House of Representatives staff, spoke to the committee about the need for accurate risk assessments and exposure measures. Henry Habicht, Michael Shapiro, Robert Kellum, and William Farland of EPA discussed where EPA was in risk assessment and how it got there. Their briefings enabled the committee to get off to a quick start in its work.

The committee was substantially helped in its activities by the strong support given it by the NRC and BEST staff: Richard D. Thomas, the project director; Deborah D. Stine, the study director; Marvin A. Schneiderman, senior scientist; Norman Grossblatt, editor, Anne M. Sprague, information specialist; Ruth E. Crossgrove, information specialist; Ruth P. Danoff, project assistant; and Shelley A. Nurse and Catherine M. Kubik, senior project assistants. Dr. Thomas had the responsibility to resolve the final technical issues and complete the report.

Finally, we must express our thanks and appreciation to the hard-working members of the committee, who struggled through long meetings, read mountains of documents, listened with interest and concern to many presentations, and then prepared what we consider to be a thoughtful, comprehensive, and balanced report that addresses many controversial issues so that all of us will have cleaner and safer air to breathe.

Kurt Isselbacher, M.D.
Chairman

Arthur Upton, M.D.
Vice Chairman

CONTENTS

EXECUTIVE SUMMARY

In recent decades, the public has become increasingly aware of seemingly innumerable reports of health threats from the environment. Myriad announcements about pesticides in food, pollutants in the air, chemical contaminants in drinking water, and hazardous-waste sites have created public concern about the chemical products and byproducts of modern industrial society. Alongside that concern is public skepticism about the reliability of scientific predictions concerning possible threats to human health. The skepticism has arisen in part because scientists disagree. But it is also apparent that many people want to understand the methods for assessing how much their exposures to chemicals threaten their health and well-being.

Many environmental issues that have risen to public prominence involve carcinogens—substances that can contribute to the development of cancer. Sometimes the decision that a substance is a carcinogen is based on evidence from workers exposed to high concentrations in the workplace, but more often it is based on evidence obtained in animals exposed to high concentrations in the laboratory. When such substances are found to occur in the general environment (even in much lower concentrations), efforts are made to determine the exposed population's risk of developing cancer, so that rational decisions can be made about the need for reducing exposure. However, scientists do not have and will not soon have reliable ways to measure carcinogenic risks to humans when exposures are small. In the absence of an ability to measure risk directly, they can offer only indirect and somewhat uncertain estimates.

Responses to these threats, often reflected in legislation and regulations, have led to reduced exposures to many pollutants. In recent years, however, concerns have arisen that the threats posed by some regulated substances might have been overstated and, conversely, that some unregulated substances might pose greater threats than originally believed. Questions have also been raised about the economic costs of controlling or eliminating emissions of chemicals that might pose extremely small risks. Debates about reducing risks and controlling costs have been fed by the lack of universal agreement among scientists about which methods are best for assessing risk to humans.

Epidemiological studies—typically, comparisons of disease rates between exposed and unexposed populations—are not sufficiently precise to find that a substance poses a carcinogenic risk to humans except when the risk is very high or involves an unusual form of cancer. For this reason, animal studies generally provide the best means of assessing potential risks to humans. However, laboratory animals are usually exposed to toxicants at concentrations much higher than those experienced by humans in the general population. It is not usually known

how similar the toxic responses in the test animals are to those in humans, and scientists do not have indisputable ways to measure or predict cancer risks associated with small exposures, such as those typically experienced by most people in the general environment.

Some hypotheses about carcinogens are qualitative. For example, biological data might suggest that any exposure to a carcinogen poses some health risk. Although some scientists disagree with that view or believe that it is not applicable to every carcinogen, its adoption provides at least a provisional answer to a vexing scientific question, namely whether people exposed to low concentrations of substances that are known to be carcinogenic at high concentrations are at *some* risk of cancer associated with the exposure. The view has dominated policy-making since the 1950s but is not always consistent with new scientific knowledge on the biological mechanisms of chemically induced cancer.

Beginning in the 1960s, toxicologists developed quantitative methods to estimate the risks associated with small exposures to carcinogens. If it were reliable, quantitative risk assessment could improve the ability of decision-makers and to some extent the public to discriminate between important and trivial threats and improve their ability to set priorities, evaluate tradeoffs among pollutants, and allocate public resources accordingly. In short, it could improve regulatory decisions that affect public health and the nation's economy.

During the 1970s and 1980s, methods of risk assessment continued to evolve, as did the underlying science. It became increasingly apparent that the process of carcinogenesis was complex, involving multiple steps and pathways. The concept that all cancer-causing chemicals act through mechanisms similar to those operative for radiation was challenged. Some chemicals were shown to alter DNA directly and hence to mimic radiation. But evidence developed that other chemicals cause cancer without directly altering or damaging DNA, for example, through hormonal pathways, by serving as mitogenic stimuli, or by causing excess cell death with compensatory cell proliferation. Biologically based and pharmacokinetic models were introduced in some cases to describe exposure-response relationships more accurately. During the same period, substantial advances were made in modeling the dispersion of airborne materials from sources to receptors and in conducting exposure assessments. Furthermore, important advances have been made in the last 10 years in understanding the basic biology of chemical toxicity. All these advances are beginning to have a major impact on the estimation of risks associated with hazardous air pollutants.

Regulation of Hazardous Air Pollutants

Before the enactment of the Clean Air Act Amendments of 1990 (1990 Amendments), Section 112 of the Clean Air Act required that the Environmental Protection Agency (EPA) set emission standards for hazardous air pollutants "to protect the public health with an ample margin of safety." In 1987, the District of Columbia Circuit Court of Appeals, in *Natural Resources Defense Council v. EPA* (824 F.2d 1146) interpreted this language to mean that EPA must first determine the emissions level that is safe—one that represents an acceptable degree of risk—and then add a margin of safety in light of the uncertainties in scientific

knowledge about the pollutant in question. The agency was permitted to consider technological feasibility in the second step but not in the first.

In response, EPA decided that it would base its regulatory decisions largely on quantitative risk assessment. The agency adopted a general policy that a lifetime cancer risk of one in 10,000 for the most exposed person might constitute acceptable risk and that the margin of safety should reduce the risk for the greatest possible number of persons to an individual lifetime risk no higher than one in 1 million (10^{-6}).

The 1990 Amendments rewrote Section 112 to place risk assessment in a key role but one secondary to technology-based regulation. As altered, Section 112 defines a list of substances as hazardous air pollutants, subject to addition or deletion by EPA. Sources that emit hazardous air pollutants will be regulated in two stages. In the first, technology-based emissions limits will be imposed. Each major source of hazardous air pollutants must meet an emission standard, to be issued by EPA, based on using the maximum achievable control technology (MACT). Smaller sources, known as area sources, must meet emissions standards based on using generally available control technology.

In the second stage, EPA must set "residual-risk standards that protect public health with an ample margin of safety if it concludes that the technology-based standards have not done so." The establishment of a residual-risk standard is required if the MACT emission standard leaves a lifetime cancer risk for the most exposed person of greater than one in a million. In actually setting the standard, though, EPA is free to continue to use its present policy of accepting higher risks. Quantitative risk assessment techniques will be relevant to this second stage of regulation, as well as to various decisions required in the first stage.

CHARGE TO THE STUDY COMMITTEE

Section 112(o) of the Act (quoted in full in Appendix M) directs the EPA to arrange for the National Academy of Sciences to:

- Review the methods used by EPA to determine the carcinogenic risk associated with exposure to hazardous air pollutants from sources subject to Section 112;
- Include in its review evaluations of the methods used for estimating the carcinogenic potency of hazardous air pollutants and for estimating human exposures to these air pollutants;
- Evaluate, to the extent practicable, risk-assessment methods for noncancer health effects for which safe thresholds might not exist.

The Academy's report must be considered by EPA in revising its present risk assessment guidelines.

CURRENT RISK-ASSESSMENT PRACTICES

Methods for estimating risk to humans exposed to toxicants have evolved steadily over the last few decades. Not until 1983, however, was the process codified in a formal way. In that year, the National Research Council released *Risk Assessment in the Federal Government: Managing the Process*. This publication, now known also as the Red Book, provided many of the definitions used throughout the environmental-health risk-assessment community today. The Red Book served as the basis for the general description of risk assessment used by the present committee.

Risk assessment entails the evaluation of information on the hazardous properties of substances, on the extent of human exposure to them, and on the characterization of the resulting risk. Risk assessment is not a single, fixed method of analysis. Rather, it is a systematic approach to organizing and analyzing scientific knowledge and information for potentially hazardous activities or for substances that might pose risks under specified conditions.

In brief, according to the Red Book, risk assessment can be divided into four steps: hazard identification, dose-response assessment, exposure assessment, and risk characterization.

- *Hazard identification* involves the determination of whether exposure to an agent can cause an increased incidence of an adverse health effect, such as cancer or birth defects, and characterization of the nature and strength of the evidence of causation.
- *Dose-response assessment* is the characterization of the relationship between exposure or dose and the incidence and severity of the adverse health effect. It includes consideration of factors that influence dose-response relationships, such as intensity and pattern of exposure and age and lifestyle variables that could affect susceptibility. It can also involve extrapolation of high-dose responses to low-dose responses and from animal responses to human responses.
- *Exposure assessment* is the determination of the intensity, frequency, and duration of actual or hypothetical exposures of humans to the agent in question. In general, concentrations of the substance can be estimated at various points from its source through the environment. An important component of exposure assessment is emission characterization, i.e., determination of the magnitude and properties of the emissions that result in exposures. This is usually accomplished by measuring and analyzing emissions, but that is not always possible. Therefore, modeling is often used instead to establish the relationship between emissions and environmental concentrations of the substance. Inputs to such a model should include data on residence and activities of the exposed population.
- *Risk characterization* combines the assessments of exposure and response under various exposure conditions to estimate the probability of specific harm to an exposed individual or population. The extent feasible, this characterization should include the distribution of risk in the population. When the distribution of risk is known, it is possible to estimate the risk to individuals who are most exposed to the substance in question.

Closely related to risk assessment is risk management, the process by which the results of

risk assessment are integrated with other information—such as political, social, economic, and engineering considerations—to arrive at decisions about the need and methods for risk reduction. The authors of the Red Book advocated a clear conceptual distinction between risk assessment and risk management, noting, for instance, that maintaining the distinction between the two would help to prevent the tailoring of risk assessments to the political feasibility of regulating the substance in question. But they also recognized that the choice of risk-assessment techniques could not be isolated from society's risk-management goals. The result should be a process that supports the risk-management decisions required by the Clean Air Act and that provides appropriate incentives for further research to reduce important uncertainties on the extent of health risks.

In 1986, EPA issued risk-assessment guidelines that were generally consistent with the Red Book recommendations. The guidelines deal with assessing risks of carcinogenicity, mutagenicity, developmental toxicity, and effects of chemical mixtures. They include default options, which are essentially policy judgments of how to accommodate uncertainties. They include various assumptions that are needed for assessing exposure and risk, such as scaling factors to be used for converting test responses in rodents to estimated responses in humans.

As risk-assessment methods have evolved and been applied with increasing frequency in federal and state regulation of hazardous substances, regulated industries, environmental organizations, and academicians have leveled a broad array of criticisms regarding the processes used by EPA. The concerns have included

- The lack of scientific data quantitatively relating chemical exposure to health risks.
- The divergence of opinion within the scientific community on the merits of the underlying scientific evidence.
- The lack of conformity among reported research results needed for risk characterization—e.g., the use of different methods for describing laboratory findings, which makes it difficult to compare the data from different laboratories and apply them in risk characterizations.
- The uncertainty of results produced by theoretical modeling, which is used in the absence of measurements.
- In response to its mandates, EPA has traditionally adopted risk assessments that for the most part incorporate conservative default options (i.e., those that are more likely to overstate than to understate human risk).
- As scientific knowledge increases, the science policy choices made by the agency and Congress should have less impact on regulatory decision-making. Better data and increased understanding of biological mechanisms should enable risk assessments that are less dependent on conservative default assumptions and more accurate as predictions of human risk.

STRATEGIES FOR RISK ASSESSMENT

The committee observed that several common themes cut across the various stages of risk assessment and arise in criticisms of each individual step. These themes are as follows:

- *Default options.* Is there a set of clear and consistent principles for modifying and departing from default options?
- *Data needs.* Is enough information available to EPA to generate risk assessments that are protective of public health and are scientifically plausible?
- *Validation.* Has the EPA made a sufficient case that its methods and models for carrying out risk assessments are consistent with current scientific information available?
- *Uncertainty.* Has EPA taken sufficient account of the need to consider, describe, and make decisions in light of the inevitable uncertainty in risk assessment?
- *Variability.* Has EPA sufficiently considered the extensive variation among individuals in their exposures to toxic substances and in their susceptibilities to cancer and other health effects?
- *Aggregation.* Is EPA appropriately addressing the possibility of interactions among pollutants in their effects on human health, and addressing the consideration of multiple exposure pathways and multiple adverse health effects?

By addressing each of those themes in each step in the risk-assessment process, EPA can improve the accuracy, precision, comprehensibility, and utility of the entire risk-assessment process in regulatory decision making.

FLEXIBILITY AND THE USE OF DEFAULT OPTIONS

EPA's risk-assessment guidelines contain a number of "default options." These options are used in the absence of convincing scientific knowledge on which of several competing models and theories is correct. The options are not rules that bind the agency; rather, they constitute guidelines from which the agency may depart when evaluating the risks posed by a specific substance. For the most part, the defaults are conservative (i.e., they represent a choice that, although scientifically plausible given existing uncertainty, is more likely to result in overestimating than underestimating human risk).

EPA has acted reasonably in electing to formulate guidelines. EPA should have principles for choosing default options and for judging when and how to depart from them. Without such principles, the purposes of the default options could be undercut. The committee has identified a number of criteria that it believes ought to be taken into account in formulating such principles: protecting the public health, ensuring scientific validity, minimizing serious errors in estimating risks, maximizing incentives for research, creating an orderly and predictable process, and fostering openness and trustworthiness. There might be additional relevant criteria.

The choice of such principles goes beyond science and inevitably involves policy choices on how to balance such criteria. After extensive discussion, the committee found that it could not reach consensus on what the principles should be or on whether it was appropriate for this committee to recommend principles. Thus, the committee decided not to do so. Appendix N contains papers by several committee members containing varied perspectives on the appropriate choice of principles. Appendix N-1 advocates the principle of "plausible conservatism" and N-2 advocates the principle of the maximum use of scientific information in selection of default options. These papers do not purport to represent the views of all committee members.

The committee did agree, though, that EPA often does not clearly articulate in its risk-assessment guidelines that a specific assumption is a default option and that EPA does not fully explain in its guidelines the basis for each default option. Moreover, EPA has not stated all the default options in the risk-assessment process or acknowledged where defaults do not exist.

EPA's practice appears to be to allow departure from a default option in a specific case when it ascertains that there is a consensus among knowledgeable scientists that the available scientific evidence justifies departure from the default option. The agency relies on its Scientific Advisory Board and other expert bodies to determine when such a consensus exists. But EPA has not articulated criteria for allowing departures.

RECOMMENDATIONS

- EPA should continue to regard the use of default options as a reasonable way to deal with uncertainty about underlying mechanisms in selecting methods and models for use in risk assessment.
- EPA should explicitly identify each use of a default option in risk assessments.
- EPA should clearly state the scientific and policy basis for each default option.
- The agency should consider attempting to give greater formality to its criteria for a departure from default options, in order to give greater guidance to the public and to lessen the possibility of ad hoc, undocumented departures from default options that would undercut the scientific credibility of the agency's risk assessments. At the same time, the agency should be aware of the undesirability of having its guidelines evolve into inflexible rules.
- EPA should continue to use the Science Advisory Board and other expert bodies. In particular, the agency should continue to make the greatest possible use of peer review, workshops, and other devices to ensure broad peer and scientific participation to guarantee that its risk-assessment decisions will have access to the best science available through a process that allows full public discussion and peer participation by the scientific community.

VALIDATION: METHODS AND MODELS

Some methods and models used in emission characterization, exposure assessment, hazard identification, and dose-response assessment are specified as default options. Others are

sometimes used as alternatives to the default options. The predictive accuracy and uncertainty of these methods and models for risk assessment are not always clearly understood or clearly explained.

A threshold model (i.e., one that assumes that exposures below some level will not cause health effects) is generally accepted for reproductive and developmental toxicants, but it is not known how accurately it predicts human risk. The fact that current evidence on some toxicants, most notably lead, does not clearly reveal a safe threshold has raised concern that the threshold model might reflect the limits of scientific knowledge, rather than the limits of safety.

EPA has worked with outside groups to design studies to refine emission estimates. However, it does not have guidelines for the use of emission estimates in risk assessment, nor does it adequately evaluate the uncertainty in the estimates.

EPA has relied on Gaussian-plume models to estimate the concentrations of hazardous pollutants to which people are exposed. These representations of airborne transport processes are approximations. EPA focuses primarily on stationary outdoor emission sources of hazardous air pollutants. It does not have a specific statutory mandate to consider all sources of hazardous air pollutants, but this should not deter the agency from assessing indoor sources to provide perspective in considering risks from outdoor sources.

EPA uses the Human-Exposure Model (HEM) to evaluate exposures from stationary sources. It estimates exposures and risk for both individuals and populations. For individuals, it has traditionally used a technique to determine what is called the maximally exposed individual (MEI) by estimating the highest exposure concentration that might be found among the broad distribution of possible exposures. Estimation of the maximum exposure is based on a variety of conservative assumptions, e.g., that the MEI lives directly downwind from the pollution source for his or her entire 70-year lifetime and remains outdoors the entire time. Traditionally, only exposure by inhalation is considered. Recently, in accordance with recommendations of the agency's Science Advisory Board, EPA has begun to replace the MEI estimate with two others: the high-end exposure estimate (HEEE) and the theoretical upper-bound exposure (TUBE).

In dose-response assessment, EPA has traditionally treated almost all chemical carcinogens as inducing cancer in a similar manner, mimicking radiation. It assumes that a linearized multistage model can be used to extrapolate from epidemiological observations (e.g., occupational studies) or experimental observations at high doses in laboratory animals down to the low doses usually experienced by humans in the general population.

RECOMMENDATIONS

- EPA should more rigorously establish the predictive accuracy and uncertainty of its methods and models and the quality of data used in risk assessment.
- EPA should develop guidelines for the amount and quality of emission information required for particular risk assessments and for estimating and reporting uncertainty in

emission estimates, e.g., the predictive accuracy and uncertainty associated with each use of the HEM for exposure assessment.

- EPA should evaluate the Gaussian-plume models under realistic conditions of acceptable distances (based on population characteristics) to the site boundaries, complex terrain, poor plant dispersion characteristics, and the presence of other structures in the vicinity. Furthermore, EPA should consider incorporating such state-of-the-art techniques as stochastic-dispersion models.
- EPA should use a specific conservative mathematical technique to estimate the highest exposure likely to be encountered by an individual in the exposure group of interest.
- EPA should use bounding estimates for screening assessments to determine whether further levels of analysis are necessary. For further analyses, the committee supports EPA's development of distributions of exposures based on actual measurements, results from modeling, or both.
- EPA should continue to explore and, when scientifically appropriate, incorporate pharmacokinetic models of the link between exposure and biologically effective dose (i.e., dose reaching the target tissue).
- EPA should continue to use the linearized multistage model as a default option but should develop criteria for determining when information is sufficient to use an alternative extrapolation model.
- EPA should develop biologically based quantitative methods for assessing the incidence and likelihood of noncancer effects in human populations resulting from chemical exposure. These methods should incorporate information on mechanisms of action and differences in susceptibility among populations and individuals that could affect risk.
- EPA should continue to use as one of its risk-characterization metrics, upper-bound potency estimates of the probability of developing cancer due to lifetime exposure. Whenever possible, this metric should be supplemented with other descriptions of cancer potency that might more adequately reflect the uncertainty associated with the estimates.

Priority-Setting and Data Needs

EPA does not have the exposure and toxicity data needed to establish the health risks associated with all 189 chemicals identified as hazardous air pollutants in the 1990 Amendments. Furthermore, EPA has not defined how it will determine the types, quantities, and quality of data that are needed to assess the risks posed by facilities that emit any of those 189 chemicals or how it will determine when site-specific emission and exposure data are needed.

Recommendations

- EPA should compile an inventory of the chemical, toxicological, clinical, and epidemiological literature on each of the 189 chemicals identified in the 1990 Amendments.

- EPA should screen the 189 chemicals to establish priorities according to procedures described by the committee for assessing health risks, identify data gaps, and develop incentives to expedite the generation of data by other government agencies (e.g., the National Toxicology Program, the Agency for Toxic Substances and Disease Registry, and state agencies), industry, and academe.
- In addition to stationary sources of hazardous air pollutants, EPA should consider mobile and indoor sources; the latter might be even more important than outdoor sources. The agency should also explicitly consider all direct and indirect routes of exposure, such as ingestion and dermal absorption.
- EPA should develop a two-part scheme for classifying evidence on carcinogenicity that would incorporate both a simple classification and a narrative evaluation. At a minimum, both parts should include the strength (quality) of the evidence, the relevance of the animal model and results to humans, and the relevance of the experimental exposures (route, dose, timing, and duration) to those likely to be encountered by humans.

VARIABILITY

Many types of variability enter into the risk-assessment process: variability within individuals, among individuals, and among populations. Types of variability include nature and intensity of exposure and susceptibility to toxic insult related to age, lifestyle, genetic background, sex, ethnicity, and other factors.

Interindividual variability is not generally considered in EPA's cancer risk assessments. The agency's consideration of variability has been limited largely to noncarcinogenic effects, such as asthmatic responses to sulfur dioxide exposure. Analyses of such variability usually form the basis of decisions about whether to protect both the general population and sensitive individuals.

RECOMMENDATIONS

- Federal agencies should sponsor molecular, epidemiological, and other types of research to examine the causes and extent of interindividual variability in susceptibility to cancer and the possible correlations between susceptibility and such covariates as age, race, ethnicity, and sex. Results should be used to refine estimates of risks to individuals and the general population.
- EPA should adopt a default assumption for differences in susceptibility among humans in estimating individual risks.
- EPA should increase its efforts to validate or improve the default assumption that humans on average have the same susceptibility as humans in epidemiological studies, the most sensitive animals tested, or both.

- EPA's guidelines should clearly state a default assumption of nonthreshold, low-dose linearity for genetic effects on which adequate data might exist (e.g., data on chromosomal aberrations or dominant or X-linked mutations) so that a reasonable quantitative estimate of genetic risk to the first and later generations can be made for environmental chemical exposure.
- The distinction between uncertainty and individual variability should be maintained rigorously in each component of risk assessment.
- EPA should assess risks to infants and children whenever it appears that their risks might be greater than those of adults.

Uncertainty

There are numerous gaps in scientific knowledge regarding hazardous air pollutants. Hence, there are many uncertainties in risk assessment. When the uncertainty concerns the magnitude of a quantity that can be measured or inferred from assumptions, such as exposure, the uncertainty can be quantified. Other uncertainties pertain to the models being used. These stem from a lack of knowledge needed to determine which scientific theory is correct for a given chemical and population at risk and thus which assumptions should be used to derive estimates. Such uncertainties cannot be quantified on the basis of data.

The upperbound point estimate of risk typically computed by EPA does not convey the degree of uncertainty in the estimate. Thus, decision-makers do not know the extent of conservatism, if any, that is provided in the risk estimate.

Formal uncertainty analysis can help to inform EPA and the public about the extent of conservatism that is embedded in the default assumptions. Uncertainty analysis is especially useful in identifying where additional research is likely to resolve major uncertainties.

Uncertainty analysis should be an iterative process, moving from the identification of generic uncertainties to more refined analyses for chemical-specific or industrial plant-specific uncertainties. The additional resources needed to conduct the more specific analyses can be justified when the health or economic impacts of the regulatory decision are large and when further research is likely to change the decision.

Recommendations

- EPA should conduct formal uncertainty analyses, which can show where additional research might resolve major uncertainties and where it might not.
- EPA should consider in its risk assessments the limits of scientific knowledge, the remaining uncertainties, and the desire to identify errors of either overestimation or underestimation.
- EPA should develop guidelines for quantifying and communicating uncertainty (e.g., for models and data sets) as it occurs into each step in the risk-assessment process.

- Despite the advantages of developing consistent risk assessments between agencies by using common assumptions (e.g., replacing surface area with body weight to the 0.75 power), EPA should indicate other methods, if any, that might be more accurate.
- When ranking risks, EPA should consider the uncertainties in each estimate, rather than ranking solely on the basis of point estimate value. Risk managers should not be given only a single number or range of numbers. Rather, they should be given risk characterizations that are as robust (i.e., complete and accurate) as can be feasibly developed.

AGGREGATION

Typically, people at risk are exposed to a mixture of chemicals, each of which might be associated with an increased probability of one or more health effects. In such cases, data are often available on only one of the adverse effects (e.g., cancer) associated with each chemical. At issue is how best to characterize and estimate the potential aggregate risk posed by exposure to a mixture of toxic chemicals. Furthermore, emitted substances might be carried to and deposited on other media, such as water and soil, and cause people to be exposed via routes other than inhalation, e.g., by dermal absorption or ingestion. EPA has not yet indicated whether it will consider multiple exposure routes for regulation under the 1990 Amendments, although it has done so in other regulatory contexts, e.g., under Superfund.

EPA adds the risks related to each chemical in a mixture in developing its risk estimate. This is generally considered appropriate when the only risk characterization needed is a point estimate for use in screening. When a more comprehensive uncertainty characterization is desired, EPA should adopt the following recommendations.

RECOMMENDATIONS

- EPA should consider using appropriate statistical (e.g., Monte Carlo) procedures to aggregate cancer risks from exposure to multiple compounds.
- In the analysis of animal bioassay data on the occurrence of multiple tumor types, the cancer potencies should be estimated for each relevant tumor type that is related to exposure, and the individual potencies should be summed for those tumors.
- Quantitative uncertainty characterizations conducted by EPA should appropriately reflect the difference between uncertainty and interindividual variability.

COMMUNICATING RISK

Certain expressions of probability are subjective, whether qualitative (e.g., that a threshold might exist) or quantitative (e.g., that there is a 90% probability that a threshold exists). Although quantitative probabilities could be useful in conveying the judgments of individual

scientists to risk managers and to the public, the process of assessing probabilities is difficult. Because substantial disagreement and misunderstanding concerning the reliability of single numbers or even a range of numbers can occur, the basis for the numbers should be set forth clearly and in detail.

RECOMMENDATION

- Risk managers should be given characterizations of risk that are both qualitative and quantitative, i.e., both descriptive and mathematical.

AN ITERATIVE APPROACH

Resources and data are not sufficient to perform a full-scale risk assessment on each of the 189 chemicals listed as hazardous air pollutants in the 1990 Amendments, and in many cases no such assessment is needed. After MACT is applied, it is likely that some of the chemicals will pose only de minimis risk (a risk of adverse health effects of one in a million or less). For these reasons, the committee believes that EPA should undertake an iterative approach to risk assessment. An iterative approach would start with relatively inexpensive screening techniques—such as a simple, conservative transport model—and then for chemicals suspected of exceeding de minimis risk move on to more resource-intensive levels of data-gathering, model construction, and model application. To guard against serious underestimations of risk, screening techniques must err on the side of caution when there is uncertainty about model assumptions or parameter values.

RECOMMENDATIONS

- EPA should develop the ability to conduct iterative risk assessments that would allow improvements to be made in the estimates until (1) the risk is below the applicable decision-making level, (2) further improvements in the scientific knowledge would not significantly change the risk estimate, or (3) EPA, the emission source, or the public determines that the stakes are not high enough to warrant further analysis. Iterative risk assessments would also identify needs for further research and thus provide incentives for regulated parties to undertake research without the need for costly, case-by-case evaluations of each individual chemical. Iteration can improve the scientific basis of risk-assessment decisions while responding to risk-management concerns about such matters as the level of protection and resource constraints.

OVERALL CONCLUSIONS AND RECOMMENDATIONS

The committee's findings are dominated by four central themes:

- Because of limitations on time, resources, scientific knowledge, and available data, EPA should generally retain its conservative, default-based approach to risk assessment for screening analysis in standard-setting; however, several corrective actions are needed to make this approach more effective.
- EPA should develop and use an iterative approach to risk assessment. This will lead to an improved understanding of the relationship between risk assessment and risk management and an appropriate blending of the two.
- The iterative approach proposed by the committee allows for improvements in the default-based approach by improving both models and the data used in analysis. For this approach to work properly, however, EPA needs to provide justification for its current defaults and establish a procedure that permits departures from the default options.
- When EPA reports estimates of risk to decision-makers and the public, it should present not only point estimates of risk, but also the sources and magnitudes of uncertainty associated with these estimates.

Risk assessment is a set of tools, not an end in itself. The limited resources available should be spent to generate information that helps risk managers to choose the best possible course of action among the available options.

1
INTRODUCTION

In recent decades, there have been seemingly innumerable reports of health threats from the environment. Myriad announcements about pesticides in food, pollutants in the air, chemical contaminants in drinking water, and hazardous-waste sites have created public concern about the chemical products and byproducts of modern industrial society. Alongside that concern exists skepticism about many of the possible threats to human health. The skepticism has arisen in part because scientists disagree. But it is also apparent that most people want to understand whether and how much their exposures to chemicals threaten their health and well-being.

Many environmental issues that have risen to public prominence involve carcinogens—substances that can contribute to the development of cancer. Sometimes the decision that a substance is a carcinogen is based on evidence from workers exposed to high concentrations in the workplace, but more often it is based on evidence obtained in animals exposed to high concentrations in the laboratory. When such substances are found to occur in the general environment (even in much lower concentrations), efforts are made to determine the exposed population's risk of developing cancer, so that rational decisions can be made about the need for reducing exposure. However, scientists do not have and will not soon have reliable ways to measure carcinogenic risks when exposures are small. In the absence of an ability to measure risk directly, they can offer only indirect and somewhat uncertain estimates.

Some hypotheses about carcinogens are qualitative. For example, biological data suggests that any exposure to a carcinogen may pose some health risk. Although some scientists disagree with that view or believe that it is not applicable to every carcinogen, its adoption provides at least a provisional answer to a vexing scientific question, namely whether people exposed to low concentrations of substances that are known to be carcinogenic at high concentrations are at *some* risk of cancer associated with the exposure. That view has been prominent since the 1950s and has guided much decision-making. For example, the "Delaney clause" of the Food Additive Amendments of 1958 stipulated that no additive that was found to be carcinogenic could be allowed in the food supply, on the grounds that it was not possible to specify a safe human exposure to such an agent. The policies that have flowed from regulations like the Delaney clause involve, where possible, absolute prohibition of exposures to carcinogens, but more commonly, reductions of exposures to the "lowest technically feasible level."

A qualitative response to the question of carcinogenic risk is still viewed by many scientists

to be the best that can now be offered, even in the face of impressive scientific advances in understanding chemical carcinogenesis. Nonetheless, it is increasingly recognized that division of the binary division of the world of chemicals into carcinogens and non-carcinogens is overly simplistic and does not provide an adequate basis for regulatory decision-making. Beginning in the 1960s and coming to full force in the 1970s, some scientists have attempted to offer more useful, quantitative information about the risks of low exposures to carcinogens. Quantitative risk assessment is attractive because, at least ideally, it allows decision-makers and the public to discriminate between important and trivial threats (thus going beyond qualitative findings that there is some risk, however small).

The results of risk assessments are important in influencing important regulatory decisions that affect both the nation's economy and public health. They influence decision-makers as they attempt to balance the view that emission of hazardous air pollutants should be minimized or even eliminated, versus the view that meeting stringent control standards might cause other problems unacceptable to society. Accurate risk assessments are also needed to determine whether public health protection is adequate.

CHARGE TO THE COMMITTEE

The charge to the committee comes from Section 112(o) of the Clean Air Act, as added by the Clean Air Act Amendments of 1990, which requires EPA to enter into a contract with the National Research Council (NRC). NRC created the Committee on Risk Assessment of Hazardous Air Pollutants in the Board on Environmental Studies and Toxicology. Its charge is summarized as follows:

1. Review the risk assessment methods used by EPA (Environmental Protection Agency).
2. Evaluate methods used for estimating the carcinogenic potencies of hazardous air pollutants.
3. Evaluate methods used for estimating human exposures to hazardous air pollutants.
4. To the extent practicable, evaluate risk-assessment methods for noncancer health effects for which safe thresholds might not exist.
5. Indicate revisions needed in EPA's risk-assessment guidelines.

The specific congressional language is provided in Appendix M. Section 112(o) requires that if EPA decides not to comply with all of the report's recommendations and the Science Advisory Board's views of the report, it must provide a detailed explanation in the *Federal Register* of the reasons that any of the recommendations in the report are not implemented.

In its charge to EPA, Congress assigned NRC the task of evaluating whether EPA's risk-assessment methods express in a scientifically supportable way the risks posed by a substance. We therefore ask whether EPA's methods are consistent with current scientific knowledge. We also ask whether EPA's methods give policy-makers and the public the information they need to make judgments about risk management. Such methods should be logical and

consistent and should, in particular, reveal the inevitable uncertainties in the underlying science.

We make no judgment regarding the appropriate risk-management decision, e.g., the extent to which society should control hazardous air pollutants. Such decisions ultimately hinge on nonscientific issues; for instance, the extent of risk from hazardous air pollutants that society is willing to accept in return for other benefits. Such issues involve not only science or science-policy judgments, but also matters of value on which scientists cannot purport to have any special insight. Such issues are therefore ultimately the province of policy-makers and the public.

It was precisely for this reason, we believe, that Congress specified in the Clean Air Act Amendments of 1990 that this committee is to undertake an investigation of EPA's risk-assessment methods, rather than of the validity of EPA's regulatory decisions. We have therefore refrained from addressing such risk management issues. We do, however, note that risk assessment and risk management are integrally related. As we explain later, Congress has generally directed EPA to be protective of health ("conservative" in the lexicon of public health) in its risk-management decisions. It is therefore essential for us to appraise whether EPA's risk-assessment methods are capable of supporting a policy of protective public-health regulation.

In addition, in its charge to EPA, Congress indicated that noncancer effects should be addressed to the extent feasible, but time constraints reduced the committee's ability to focus fully on this issue.

Section 303 of the 1990 Amendments created the Risk Assessment and Management Commission, part of whose charge is to examine risk-management policy issues. Specific subjects that the commission is to address are

- The report of the NRC committee.
- The use and limitations of risk assessment in establishing emission or effluent standards, ambient standards, exposure standards, acceptable concentrations, tolerances, or the environmental criteria for hazardous substances that present a risk of carcinogenic or other chronic health effects and the suitability of risk assessment for such purposes.
- The most appropriate methods for measuring and describing cancer risks or risks of other chronic health effects associated with exposure to hazardous substances.
- Methods to reflect uncertainties in measurement and estimation techniques, the existence of synergistic or antagonistic effects among hazardous substances, the accuracy of extrapolating animal-exposure data to human health risks, and the existence of unquantified direct or indirect effects on human health in risk-assessment studies.
- Risk-management policy issues, including the use of lifetime cancer risks to the people most exposed, the incidence of cancer, the cost and technical feasibility of exposure-reduction measures, and the use of site-specific actual exposure information in setting emission standards and other limitations applicable to sources of exposure to hazardous substances.
- The degree to which it is possible or desirable to develop a consistent risk-assessment method, or a consistent standard of acceptable risk among various federal programs.

Besides the Academy's report and the activities of the commission, both EPA and the Surgeon General are to evaluate the methods for evaluating health risks, the significance of residual risks, uncertainties associated with this analysis, and recommend legislative changes.

As a result, the committee highlights here some important and controversial subjects in risk assessment and management that it felt were beyond its charge.

1. *The use of a specific individual lifetime cancer risk number (e.g., 10^{-4} or 10^{-6}) as a target for risk regulations.* The committee notes that Congress has set a standard for considering regulatory decisions. We note that such a number should be tied to a method and that uncertainty will always surround such estimates.

2. *The use of comparative risk analysis for the allocation of resources to minimize health impacts.* Congress decides how much of the country's economic and social resources should be spent on reducing threats to public health and how to allocate resources among the many threats present in our daily lives.

3. *The relative risk associated with synthetic or industrial byproducts versus natural chemicals.* A recent study (Gold et al., 1992) contends that natural chemicals make up the vast bulk of chemicals to which humans are exposed, that natural chemicals are not much different from synthetic chemicals in their toxicology, and that about half the natural chemicals tested in chronic studies in rats and mice are carcinogens. The implication is that humans are likely to be exposed to a large background of rodent carcinogens as defined by high-dose testing. Some believe that this has implications for the amount of resources currently devoted to the study and control of synthetic chemicals. However, other studies (e.g., Perera and Bofetta, 1988) question the scientific underpinnings of these conclusions. The issue of the degree to which natural versus synthetic chemicals should be regulated is a policy issue that we cannot address. The scientific aspects of the issue will be discussed in a forthcoming NRC report on the relative risks of natural carcinogens. It is important to note that the present study focuses on airborne hazardous air pollutants and that, although some natural carcinogens are in food and water, there is little evidence of their widespread presence in air.

4. *The setting of relative policy priorities regarding the regulation of all sources of hazardous air pollutants.* The focus of Section 112 is on stationary sources of hazardous air pollutants; therefore, it was not within the charge of this committee to conduct an analysis of all sources of hazardous air pollutants and recommend which ones should be regulated and which should not. Congress already determined the extent to which it wanted to do that in the 1990 Amendments. Therefore, although the committee points out later in the report the potential impact of indoor versus outdoor pollutants, it is beyond our charge to go further and say whether, when, and how to take action on nonstationary and indoor sources of hazardous air pollutants.

5. *The uncertainty in engineering and economic assumptions.* There is, of course, uncertainty in the engineering and economic assumptions leading to EPA's estimates of the impact on industry of a regulation mandating specified magnitudes of risk. However, the committee was asked only to address EPA's implementation of risk assessment relative to public health, not the economic consequences of such regulation.

6. *The extent to which chemicals should be on or off the list of chemicals in the 1990 Amendments.* Although this report discusses how to set priorities for the collection and analysis of chemicals on the list, it is a policy judgment as to whether these chemicals, once ranked, should be included on such a list. (That does not imply that outside review of the list is not appropriate.)

7. *The presentation of uncertainty in the context of background risk.* Although this committee does discuss the issue of presentation of uncertainty, it was beyond its charge to indicate the extent to which it was appropriate to place the 1990 Amendments or other legislation in the context of all societal risk. Risk communication is complicated and involves such issues as involuntary versus voluntary risks, costs, benefits, and values, both individual and societal.

CONCEPTUAL FRAMEWORK OF THE REPORT

This report is aimed at a multidisciplinary audience with different levels of technical understanding. In discussing the many controversial aspects of risk assessment, the committee decided to address three categories of issues:

- Background of risk assessment and current practice at EPA. We organize this section (Chapters 2-5) via the old Red Book four-step paradigm.
- Specific concerns in risk assessment, such as the use of defaults and extrapolations. For example, is EPA justified in assuming, in the absence of contrary evidence, that the linearized multistage model should be used in determining the dose-response relationship for carcinogens?
- Cross-cutting issues that affect all parts of risk assessment. For example, how should uncertainty be handled? How should the accuracy of a model be evaluated?
- Implementation issues related to Section 112 of the 1990 Amendments. For example, how should EPA accommodate the tension between the goals of providing stability in its process and staying abreast of changing scientific knowledge?

The report addresses each type of issue. Our categorization of the issues reflects the analytical framework used by the committee and influences the structure of its recommendations. Although that might lead to some repetition, the committee feels that a degree of repetition is desirable because of the need to address audiences with different levels of knowledge.

The committee attempted to address the specific issues that arise from the uses of risk assessment under Section 112 of the Clean Air Act, which mandates the regulation of hazardous air pollutants. As amended in 1990, Section 112 de-emphasizes risk assessment in the initial phase of regulation, in which EPA is to establish "technology-based" standards for categories of sources that emit hazardous air pollutants. Risk assessment's main role will be in the second phase of regulation, in which EPA must determine whether residual risk (the risk presented by the emissions remaining after compliance with technology-based standards) should be further reduced. Risk assessment will also be used in several other ways (e.g., to

determine whether an entire source category may be exempted from technology-based standards on the grounds that no source in the category creates more than a one-in-a-million lifetime risk of cancer for the most exposed person).

The appendixes to the report include EPA's responses to questions from the committee and some important EPA documents not readily available. Risk assessment is an ever-changing process, and these documents illustrate its status within EPA during the time when the committee is making its recommendations. Two documents were also prepared by some committee members to reflect the committee's inability to reach consensus on how EPA should choose and refine its "default options" for conducting risk assessments when basic scientific mechanisms are unknown. One view espouses a principle of "plausible conservatism," while the other advocates "making full use of science."

Part I
CURRENT APPROACHES TO RISK ASSESSMENT

The first part of the report examines the background and current practices of risk assessment consistent with the paradigm first codified in the 1983 NRC report *Risk Assessment in the Federal Government: Managing the Process*, often known as the Red Book (See Figure I-1). Chapter 2 of this report discusses the historical, social, and regulatory contexts of quantitative risk assessment. Chapters 3, 4, and 5 describe the Environmental Protection Agency's approach in applying the Red Book paradigm for risk assessment. As shown in Figure I-2, assessing human-health risks associated with a pollutant requires analysis of three elements: the *source* of the pollutant, the transport of the pollutant into the *environment* (air, water, land, and food), and the intake of the pollutant by *people* who might suffer adverse health effects either soon after exposure or later. Scientists and engineers take four basic interrelated steps to evaluate the potential health impact on people who are exposed to a hazardous air pollutant: emission characterization, exposure assessment, toxicity assessment, and risk characterization. In emission characterization, the chemical's identity and the magnitude of its emissions are determined. Exposure assessment includes how the pollutant moves from a source through the environment (transport) until it is converted to other substances (fate) or comes into contact with humans. In assessment of toxicity, the specific forms of toxicity that can be caused by the pollutant and the conditions under which these forms of toxicity might appear in exposed humans are evaluated. In risk characterization, the results of the analysis are described. These steps are described in detail in Chapters 3, 4, and 5.

The increase in the sophistication of the field of risk assessment since the Red Book requires risk assessors to have the ability to recognize and address fully such cross-cutting issues as uncertainty, variability, and aggregation, in addition to having a more overarching view of the practice of risk assessment. Therefore, the committee supplements the Red Book paradigm with a second approach—one that is less fragmented (and hence more holistic), less linear and more interactive, and, most important, one organized not according to discipline or function, but according to the recurring conceptual issues that cut across all the stages of risk assessment. These cross-cutting issues are described in Part II of this report.

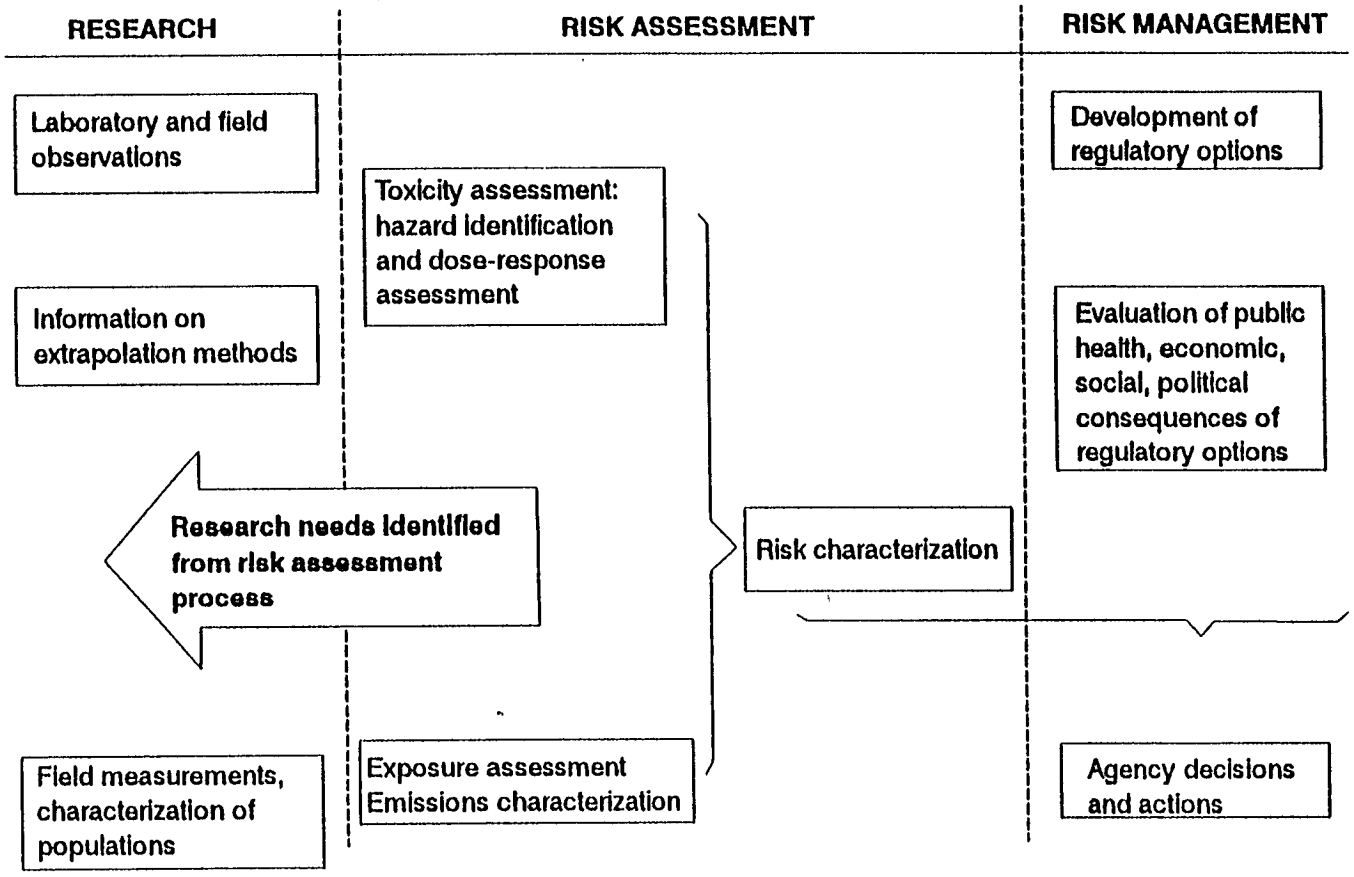


Figure I-1. NAS/NRC risk assessment/management paradigm. Source: Adapted from NRC, 1983a.

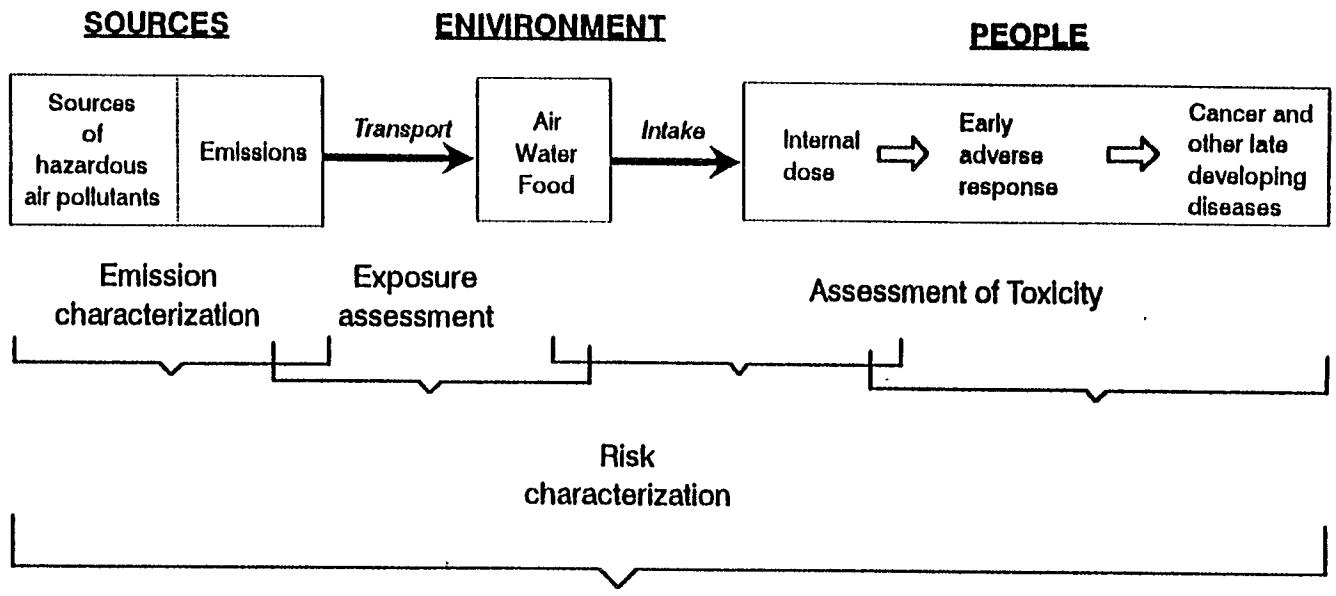


Figure I-2. Relationships in assessing human health risks of exposure to hazardous air pollutants. Source: Adapted from NRC, 1983a.

2

RISK ASSESSMENT AND ITS SOCIAL AND REGULATORY CONTEXTS

This chapter provides an overview of the origins and uses of quantitative risk assessment and the problems associated with it. Historical perspective is offered to aid understanding of how a method infused with so much uncertainty has still come to be seen by many as useful. Some attention is devoted to the important questions of how risk assessment has been used in decision-making and whether its use has improved decisions. The issues of public acceptance of the method and the degree to which decisions based on it are seen to provide adequate protection of the public health are also addressed. This chapter lists the major criticisms of risk assessment and the ways in which its results have been used, thus providing the justification for the selection of issues discussed in the succeeding chapters.

GENERAL CONCEPTS

This section briefly discusses some basic definitions and concepts concerning human-health risk assessment, its content, and its relationships to research and to decision-making. The definitions and concepts were first systematically formulated by a National Research Council committee in a report issued in 1983, *Risk Assessment in the Federal Government: Managing the Process*. The Red Book had a major influence on the practice of risk assessment and will be discussed extensively in this section of the report.

WHAT IS RISK ASSESSMENT?

Human-health risk assessment entails the evaluation of scientific information on the hazardous properties of environmental agents and on the extent of human exposure to those agents. The product of the evaluation is a statement regarding the probability that populations so exposed will be harmed, and to what degree. The probability may be expressed quantitatively or in relatively qualitative ways. There are other types of risk assessment that use similar processes but are outside the scope of this report, e.g., the risk assessment of the relative safety of a bridge.

Chemical hazards come in many forms. Some substances are radioactive, some explosive,

some highly flammable. The particular hazard of concern here is chemical toxicity, including but not limited to carcinogenicity. Risk assessments can be carried out for any form of chemical toxicity. Risk assessment can be qualitative or quantitative. Many of the issues covered in this report concern quantitative expressions of risk.

How Is Risk Assessment Conducted?

The 1983 NRC report described a four-step analytic process for human-health risk assessment. A substance leaves a source (e.g., an industrial facility), moves through an environmental medium (e.g., the air), and results in an exposure (people breathe the air containing the chemical). The exposure creates a dose in the exposed people (the amount of the chemical entering the body, which may be expressed in any of several ways), and the magnitude, duration, and timing of the dose determine the extent to which the toxic properties of the chemical are realized in exposed people (the risk). This model is captured in the following analytic steps:

Step 1: Hazard Identification entails identification of the contaminants that are suspected to pose health hazards, quantification of the concentrations at which they are present in the environment, a description of the specific forms of toxicity (neurotoxicity, carcinogenicity, etc.) that can be caused by the contaminants of concern, and an evaluation of the conditions under which these forms of toxicity might be expressed in exposed humans. Information for this step is typically derived from environmental monitoring data and from epidemiologic and animal studies and other types of experimental work. This step is common to qualitative and quantitative risk assessment.

Step 2: Dose-Response Assessment entails a further evaluation of the conditions under which the toxic properties of a chemical might be manifested in exposed people, with particular emphasis on the quantitative relation between the dose and the toxic response. The development of this relationship may involve the use of mathematical models. This step may include an assessment of variations in response, for example, differences in susceptibility between young and old people.

Step 3: Exposure Assessment involves specifying the population that might be exposed to the agent of concern, identifying the routes through which exposure can occur, and estimating the magnitude, duration, and timing of the doses that people might receive as a result of their exposure.

Step 4: Risk Characterization involves integration of information from the first three steps to develop a qualitative or quantitative estimate of the likelihood that any of the hazards associated with the agent of concern will be realized in exposed people. This is the step in

which risk-assessment results are expressed. Risk characterization should also include a full discussion of the uncertainties associated with the estimates of risk.

Not every risk assessment encompasses all four steps. Risk assessment sometimes consists only of a hazard assessment designed to evaluate the potential of a substance to cause human health effects. Regulators sometimes take the additional step of ranking the potency of a number of chemicals—what is known as hazard ranking. Sometimes potency information is combined with exposure data to produce a risk ranking. These techniques all use some, but not all, of the four steps of the quantitative risk-assessment process.

Much of this report is devoted to the technical contents of the four steps of the process, because therein lie the issues that affect the reliability, utility, and credibility of risk-assessment outcomes. One important feature of those steps, however, needs to be emphasized here.

The 1983 NRC committee recognized that completion of the four steps rests on many judgments for which a scientific consensus has not been established. Risk assessors might be faced with several scientifically plausible approaches (e.g., choosing the most reliable dose-response model for extrapolation beyond the range of observable effects) with no definitive basis for distinguishing among them. The earlier committee pointed out that selection of a particular approach under such circumstances involves what it called a *science-policy* choice. Science-policy choices are distinct from the policy choices associated with ultimate decision-making, as will be seen below. The science-policy choices that regulatory agencies make in carrying out risk assessments have considerable influence on the results and are the focus of much that follows in this report.

What is the Relationship Between Risk Assessment and Research?

Although the conduct of a risk assessment involves research of a kind, it is primarily a process of gathering and evaluating extant data and imposing science-policy choices. Risk assessment draws on research in epidemiology, toxicology, statistics, pathology, molecular biology, biochemistry, analytical chemistry, exposure modeling, dosimetry, and other disciplines; to the extent that it attempts to capture and take into account uncertainties, it also draws on the research efforts of decision analysts.

Risk assessment, at least in theory, can influence research directions. Because, at its best, risk assessment provides a highly organized profile of the current state of knowledge of particular issues and systematically elucidates scientific uncertainties, it can provide valuable guidance to research scientists regarding the types of data that can most effectively improve understanding. Little effort seems to have been made to use risk assessments in this way, although the Office of Technology Assessment has recently completed a study that describes the role of risk assessment in guiding research (OTA, 1993).

What is the Relationship Between Risk Assessment and Regulatory Decision-Making?

Risk management is the term used to describe the process by which risk-assessment results are integrated with other information to make decisions about the need for, method of, and extent of risk reduction. Policy considerations derived largely from statutory requirements dictate the extent to which risk information is used in decision-making and the extent to which other factors—such as technical feasibility, cost, and offsetting benefits—play a role.

Some statutes seem not to permit risk-assessment results to play a substantial role; they stress reductions of exposure to the "lowest technically feasible level" and usually require the best available technology. Proponents of such technology-based approaches often argue that they facilitate more rapid regulatory action and are especially suitable for making large and relatively inexpensive "first-cut" emission reductions. Proponents of quantitative risk assessment argue that such approaches are blind to the possibility that the risks remaining after application of such technology might still be unreasonably large or, in other situations, that they have been pushed to unnecessarily low values. As amended in 1990, Section 112 of the Clean Air Act gives quantitative risk-assessment results a secondary but still important role relative to technology-based controls.

What Is a Default Option?

EPA's guidelines set forth "default options." These are generic approaches, based on general scientific knowledge and policy judgment, that are applied to various elements of the risk assessment process when specific scientific information is not available. For instance, ambient doses of contaminants in humans are generally far lower than the doses that produce tumors in animals in controlled studies. The guidelines advise that, in assessing the magnitude of cancer risk to humans from low doses of a chemical based on the results of a high-dose experiment, "in the absence of adequate information to the contrary, the linearized multistage procedure will be employed" (EPA, 1986a, 1987a); that is, cancer risk in humans exposed to low doses will be estimated mathematically by using high-dose data and a curve-fitting procedure to extrapolate to low doses. Departure from the guideline is allowed if there is "adequate evidence" that the mechanism through which the substance is carcinogenic is more consistent with a different model; for instance, that there is a threshold below which a substance will not cause a risk. Thus, the guideline amounts to a "default" that guides a decision-maker in the absence of evidence to the contrary; in effect, it assigns the burden of persuasion to those wishing to show that the linearized multistage procedure should not be used. Similar guidelines cover such important issues as the calculation of effective dose, the consideration of benign tumors, and the procedure for scaling animal-test results to estimates of potency in humans. In the absence of information on some critical point in a risk assessment, default procedures seem essential. The question, then, is not whether to use defaults,

but which defaults are most appropriate for a specific task and when it is appropriate to use an alternative to a default.

HISTORICAL ROOTS

It is helpful to provide a brief historical perspective on the origins and evolution of risk assessment, so that some of the reasons that led to the use of the technique can be seen. The review is divided into two main parts, with an intervening section devoted to the NRC study of 1983 that was so influential in the developments of the last decade.

EARLY EFFORTS TO ESTABLISH SAFE LIMITS OF EXPOSURE TO TOXIC SUBSTANCES

About 50 years ago, toxicologists began to study the problem of establishing limits on exposures to hazardous substances that would protect human health. The early efforts began in the 1940s in connection with concerns about occupational exposures to chemicals and about residues of pesticides in foods. Toxicologists were guided by the principle that all substances could become harmful under some conditions of exposure—when the so-called threshold dose was exceeded—but that human health could be protected as long as those exposure conditions were avoided. Threshold doses were recognized to vary widely among chemicals, but as long as human exposures were limited to subthreshold doses, no injury to health would be expected. The threshold hypothesis thus involved rejection of the simplistic view that the world is divided into toxic and nontoxic substances and acceptance of the principle that, for all chemicals, there were ranges of exposure that were toxic and ranges that were not. The threshold hypothesis was based on both empirical observations and basic concepts of biology—that every organism, including the human, has the capacity to adapt to or otherwise tolerate some exposure to any substance and that the harmful effects of a substance would become manifest only when exposure exceeded that capacity. Even at that early stage, there were questions about whether carcinogens always had thresholds, but otherwise the threshold concept became widely accepted.

Although there was widespread acceptance of the threshold hypothesis (except among scientists working in genetics and in chemical carcinogenesis) (NRC, 1986), it was not apparent how the threshold dose was to be estimated for a large and diverse human population whose members have different thresholds of susceptibility. Experts in occupational health tended to rely heavily on observations of short-term toxicity in highly exposed workers and established acceptable exposure limits (the most prominent of which were the so-called threshold limit values, TLVs, first published by the American Conference of Governmental Industrial Hygienists in the 1950s) that were below the exposures that produced observable toxic effects. In the early 1950s, two Food and Drug Administration (FDA) scientists, O.G. Fitzhugh and A. Lehman, proposed a procedure for setting acceptable limits, which became

known as acceptable daily intakes (ADIs), for dietary pesticide residues and food additives. Their procedure was based on the threshold hypothesis and first involved identification of a chemical's no-observed-effect level (NOEL) from the set of chronic animal-toxicity data in which the animals responded to the lowest dose tested—the "most sensitive" indication of the chemical's toxicity. Several response levels are characterized by acronyms. The first is the "no-observed-effect-level," NOEL. Earlier this was called the no-*observable*-effect level. *Observable* was changed to *observed* to be more in keeping with actual data ("observed"), rather than a rather vague potential "observable," which might be related to the size and sensitivity of the experiment. What is not observable in a small experiment might be easily observed in a large experiment. The word *adverse* was added to NOEL, making it NOAEL and making it clearer that *adverse* effects were of concern. The LOEL and LOAEL have a similar genesis and currently refer to the "lowest-observed-adverse-effect level"—the lowest dose at which an adverse effect was seen.

Fitzhugh and Lehman cited data suggesting that "average" human sensitivities might be up to 10 times those of laboratory animals and that some members of a large and diverse human population might be up to 10 times more sensitive than the "average" person. Thus came into use the safety factor of 100. The experimental NOEL was divided by 100 to arrive at a chemical-specific ADI. If human exposure was limited to daily amounts less than the ADI, then no toxicity was to be expected. In fact, Fitzhugh and Lehman, and later other authors and expert groups, including the World Health Organization, did not claim that an ADI arrived at in this fashion was risk-free, but only that it carried "reasonable certainty of no harm." No attempt was made to estimate the probability of harm. A variation of the safety-factor approach, often called margin of safety, is the estimate of the ratio of the NOEL to actual exposures. A judgment is made as to whether that ratio is acceptable. This margin-of-safety approach seems to be most common for substances already in general use, and in practice is often associated with lower ratios of NOEL to exposure than those based on safety factors.

The use of safety factors to establish ADIs was also recommended by various NRC committees (NRC, 1970, 1977, 1986) and adopted by the Joint Food and Agriculture Organization and World Health Organization expert committees on food additives (FAO/WHO, 1982) and pesticide residues (FAO/WHO, 1965).

Although it has since been modified in several minor ways, the basic procedure for setting limits on human exposures to chemicals in air, water, and food persists to this day. The threshold hypothesis has been criticized as inadequate to account for some toxic effects, and it has not been accepted by regulators as applicable to carcinogens, but it remains a cornerstone of other regulatory and public-health risk assessments. Section 112 of EPA's authority for regulating toxic air pollutants envisions a safety-factor approach for some kinds of risk assessment.

THE PROBLEM OF CARCINOGENS

Not only is cancer a much-feared set of diseases, but public and scientific concerns about

cancer-inducing chemicals in the environment have centered on the possibility that such substances might act through nonthreshold mechanisms; that is, that exposure to even one molecule of a carcinogen is associated with a small but non-zero increased risk of tumor induction. This possibility served as the basis for modern dose-response models, which were developed initially from observations of radiation-induced cancer. These models came into wide use and were promoted by the National Research Council's series of reports entitled *Biological Effects of Ionizing Radiation* and later incorporated into the regulatory decision-making of the Nuclear Regulatory Commission. Perhaps the earliest legislative acknowledgment of the possibility that chemical carcinogens might act in the same way came in the form of the "Delaney clause" of the Food Additive Amendments of 1958. Following the suggestions set forth by several FDA and National Cancer Institute (NCI) officials, Congress stipulated that no additive that concentrates in food during processing or is added to food during or after processing may be allowed in the food supply if it is found to be carcinogenic in animals. The basis for the Delaney clause was that it is not possible to specify a safe human exposure to a carcinogen in the same sense that a safe intake of a substance acting through threshold mechanisms could be identified.

Through the 1960s and into the early 1970s, toxicologists avoided the problem of identifying "acceptable" intakes of carcinogens. Where it was possible, regulators simply prohibited introduction of carcinogens into commerce. But where banning was difficult or even infeasible— for example, for environmental contaminants that were byproducts of manufacturing and energy production—choosing a maximal permissible human exposure, and acceptance of some risk. Limits were sometimes based on some concept of technical feasibility. The problem with such a criterion for setting limits was that it provided little confidence that human health was being adequately protected or, conversely, that risks were not being forced to unnecessarily low levels. In many cases, carcinogenic pollutants were simply ignored (NRC, 1983a).

Those approaches to the problem of regulatory exposure to environmental carcinogens became problematic in the face of two trends. First, government and industrial testing for carcinogenicity began to increase rapidly during the late 1960s; during the 1970s, regulators had to begin to deal with large numbers of newly identified carcinogens that were found among the many commercial products introduced after World War II. Second, analytic chemists became able to identify carcinogens in the environment at lower and lower concentrations. It became clear during the early to middle 1970s that a systematic approach to regulating carcinogens was needed.

Several authors had published methods for quantifying low-dose risks associated with chemical carcinogen exposure in the 1960s and 1970s, and regulatory agencies—FDA and EPA in particular—began adopting some of the methods in the middle 1970s. EPA, for example, estimated low-dose risks associated with several carcinogenic pesticides and relied in part on its assessments in actions to cancel or limit their registrations. FDA began using low-dose risk estimation to deal with so-called indirect food additives and some food contaminants that proved to be carcinogenic. The Occupational Safety and Health Administration (OSHA) at first rejected the use of risk quantification as it mounted a major effort during the late 1970s to regulate occupational carcinogens, because it believed that the statute under which it operated

did not permit the use of risk assessment. But a Supreme Court decision regarding the agency's efforts to establish a permissible exposure limit for benzene caused OSHA to incorporate risk quantification (see below).

Those trends of the 1970s toward increasing the use of risk assessment in carcinogen regulation caused several regulatory agencies, working together as the Interagency Regulatory Liaison Group (IRLG), to develop and publicize a set of guidelines for the conduct of risk assessments (IRLG, 1979). The guidelines were said by the agencies to specify a common approach to risk assessment. No commitment was made by the agencies to use the methods for all possible carcinogens in all classes of regulated products, but, to the extent that an agency decided to use risk assessment, its approach would be that specified in the IRLG guidelines. The agencies also noted that the guidelines did not include an approach to what later came to be called risk management; such issues were said to remain the prerogative of the individual agencies.

The IRLG guidelines embodied several important scientific principles that originated in efforts of the WHO International Agency for Cancer Research (IARC) (IARC, 1972, 1982), NCI (Shubik, 1977), and the federal regulatory agencies (FDA, 1971; Albert et al., 1977; OSHA, 1982). Among them were principles concerning the appropriate uses of epidemiologic and animal data in identifying potential human carcinogens and the extrapolation of such data to humans. The IRLG guidelines did not explicitly incorporate the "default options" language described earlier (that came only after the 1983 NRC report), but it is clear that they do include science-policy choices (e.g., the generic adoption of a linearized, no-threshold model for carcinogen dose-response assessment).

By the early 1980s, risk assessment had begun to take on considerable importance within the regulatory agencies and to capture the attention of regulated industries. One important impetus to the development of risk-assessment techniques was the Supreme Court's decision in *Industrial Union Department, AFL-CIO v. American Petroleum Institute*, 448 U.S. 607 (1980), the "*Benzene*" decision. That decision struck down the OSHA standard for exposure to benzene in the workplace. The standard was based on OSHA's policy of trying to reduce concentrations of carcinogens in the workplace as far as technologically possible without consideration of whether existing concentrations posed a significant risk to health. There was no opinion for the majority of the Supreme Court in *Benzene,* but four justices concluded that, under the Occupational Safety and Health Act, OSHA could regulate only if it found that benzene posed a significant risk of harm. Although the plurality did not define *significant risk of harm* and stressed that the magnitude of the risk need not be determined precisely, the decision strongly signaled that some form of quantitative risk assessment was necessary as a prelude to deciding whether a risk was large enough to deserve regulation.

Under those circumstances, Congress instructed FDA to arrange for the National Research Council in 1981 to undertake a study of federal efforts to use risk assessment.

NRC Study of Risk Assessment In the Federal Government

In 1983, NRC was asked to issue recommendations regarding the scientific basis of risk assessment and the institutional arrangements under which it was being conducted and used. In particular, NRC's charge involved a close examination of the possibility that risk assessment might be conducted by a separate, centralized scientific body that would serve all the relevant agencies. It was proposed that such an arrangement might reduce the influence of policy-makers on the conduct of risk assessment, so that there would be minimal opportunities for the results of risk assessments to be manipulated to meet predetermined policy objectives.

The NRC committee drew extensively on the earlier work of EPA, FDA, OSHA, IARC, and NCI, and much of its effort was directed at a synthesis of scientific principles and concepts first elucidated by these agencies. The NRC study did not, however, recommend specific methods for the conduct of risk assessment.

The risk assessment framework and specific definitions of risk assessment and its component steps from the 1983 NRC report have been widely adopted. Many of the recommendations from the 1983 report have been implemented by EPA and other regulatory agencies. Two of the major recommendations of the committee, summarized below, are particularly relevant to this report:

- A clear conceptual distinction between risk assessment and risk management should be maintained. It is, however, not necessary—indeed, it is inadvisable—to provide for a physical separation of the two activities. (The committee rejected the proposal for the establishment of an independent scientific group that would perform risk assessments for the regulatory agencies.) Risk assessments should be undertaken with careful attention to the contexts in which those assessments will be used.
- Regulatory agencies should develop and use inference guidelines that detail the scientific basis for the conduct of risk assessment and that set forth the default options. The guidelines should be explicit about the steps of risk assessment that require such science-policy choices. The guidelines are necessary to avoid the appearance of case-by-case manipulation of assumptions to meet preset management goals. Guidelines should be flexible, however, and allow departures from defaults when data in specific cases show that a default option is not appropriate.

The NRC committee did not specify any particular methodologic approach to risk assessment, nor did it address the issue of which default options should be used by regulatory agencies. It did, however, note that provisions should be made for continuing review of the science underlying the guidelines and of the basis of the default options incorporated in them.

EVENTS AFTER RELEASE OF THE 1983 NRC REPORT

The Office of Science and Technology Policy (OSTP) brought together scientists from the regulatory agencies, the National Institutes of Health, and other federal agencies and, in 1985, issued a comprehensive review of the scientific basis of risk assessment of chemical carcinogens. The OSTP review adopted the framework for risk assessment proposed by the NRC committee and provided the individual regulatory agencies a basis for developing the type of guidelines recommended by that committee.

Alone among federal agencies, EPA adopted a set of guidelines for carcinogen risk assessment in 1986, as recommended by NRC. The EPA guidelines specify default options, note the distinction between risk assessment and risk management, and otherwise meet NRC's and OSTP's recommendations. EPA has issued guidelines for assessing risks associated with several other adverse health effects of toxic substances (without the benefit of OSTP review of the underlying science) and for the conduct of human exposure assessments. Beginning in 1984, it initiated work and published guidelines for evaluating mutagenicity, developmental toxicity, effects of chemical mixtures, and human exposure (EPA, 1986a, 1987a). It later published proposed guidelines on female reproductive risk (EPA, 1988a), male reproductive risk (EPA, 1988b), and exposure-related measurements (EPA, 1988c). Final, revised guidelines on developmental toxicity were published in 1991 (EPA, 1991a). The agency is now in the processing of issuing revised guidelines on cancer risk assessment and has issued revised guidelines for the assessment of human exposures (EPA, 1992a).

Increasing activity at the state level was first indicated by California's publication in 1985 of *Guidelines for Chemical Carcinogen Risk Assessments and Their Scientific Rationale* (CDHS, 1985). The purpose of the guidelines was "to clarify internal procedures which risk assessment staff of the California Department of Health Services will usually use to deal with certain decision points which are characteristic of most risk assessments." The authors went on to state why guidelines were thought necessary, in language consistent with earlier statements of IRLG (1979), NRC (1983a), OSTP (1985), and EPA (1987a):

> These California guidelines, while in harmony with recent federal statements on carcinogenic risk assessment, are more specific and practical. The Department of Health Services' staff believe that there are important advantages to the announcement of such flexible nonregulatory guidelines. First, the publishing of guidelines increases the likelihood of consistency in risk assessment among agencies and decreases the time spent repeatedly arguing risk assessment policy for each separate substance. Second, announcing guidelines in advance makes it clear that one is not tailoring risk assessment to justify some predetermined risk management decision. Third, specific guidelines allow the regulated community to predict what emissions, food residues, or other exposures are apt to be of public health concern. Fourth, the publication and discussion of these guidelines should make the process more understandable to risk managers who have to make decisions that depend in part on risk assessment determinations.

The NRC, OSTP, EPA, and California documents were produced during a time in which the uses of risk assessment to guide regulatory decision-making were expanding rapidly. Particularly important was EPA's adoption of risk assessment as a guide to decisions at

Superfund and other hazardous-waste sites, including those covered by the Resource Conservation and Recovery Act (RCRA).

The agency also extended the uses of risk assessment to decisions regarding pesticide residues in food, carcinogenic contaminants of drinking-water supplies, industrial emissions of carcinogens to surface waters, and industrial chemicals subject to regulation under the Toxic Substances Control Act (TSCA). Risk-management approaches varied according to the specific legal requirements applicable to the sources of carcinogen exposure, but the EPA guidelines were intended to esnure that the agency's approach to risk assessment was uniform across the various programs.

USES OF RISK ASSESSMENT IN THE REGULATION OF HAZARDOUS AIR POLLUTANTS

Section 112 of the Clean Air Act, as originally adopted in the Clean Air Act Amendments of 1970, required EPA to set emissions standards for hazardous air pollutants so as to protect public health with an "ample margin of safety." EPA was slow in carrying out that mandate, listing only eight chemicals as hazardous air pollutants in 20 years.[1] Standards were issued for only seven (there was no standard for coke ovens), and the standards that were issued covered only some of the sources that emit these pollutants. One major reason was the ambiguity of "ample margin of safety." Many commentators long thought that that term barred EPA from considering costs; EPA might well have to set a standard of zero for any pollutant for which no threshold could be defined (i.e., virtually all carcinogens).

That interpretation of the act (originally developed well before 1987), however, was unanimously rejected by the District of Columbia Circuit court in *Natural Resources Defense Council v. EPA* (824 F.2d 1146 [en banc] [D.C.Cir. 1987]). At the same time, the Court of Appeals also rejected EPA's position that it could use technologic or economic feasibility as the primary basis for standard-setting under Section 112. Instead, the court held that EPA had first to determine what concentration was "safe"—i.e., represented an acceptable degree of risk—and had then to select a margin of safety necessary to incorporate the uncertainties in scientific knowledge. In the latter step, but not the former, technological feasibility could be taken into account. In accordance with the plurality opinion in the Supreme Court's *Benzene* decision, the circuit court also held that EPA's standards did not have to eliminate all risk.

As in the *Benzene* case, the court did not define any particular method for EPA to use in determining what risks are acceptable. On remand, the agency, after taking comment on a number of possibilities, decided that it could not use any single metric as a measure of whether a risk is acceptable. Instead, it adopted a general presumption that a lifetime excess risk of

---

[1]The chemicals listed as hazardous air pollutants under the National Standards for Hazardous Air Pollutants (NESHAP) (with the date of public notice): asbestos (3/71); benzene (6/77); beryllium (3/71); coke-oven emissions (9/84); inorganic arsenic (6/80); mercury (3/71); radionuclides (12/79); and vinyl chloride (12/75).

cancer of approximately one in 10,000 (10^{-4}) for the most exposed person would constitute acceptable risk and that the margin of safety should reduce the risk for the greatest possible number of persons to an individual lifetime excess risk no higher than one in 1 million (10^{-6}). Such factors as incidence (e.g., the number of possible new cases of a disease in a population), the distribution of risks, and uncertainties would be taken into account in applying those benchmarks. The agency approach thus put primary emphasis on estimating individual lifetime risks through quantitative risk assessment.

Congress lessened the role of quantitative risk assessment for air-pollution regulation by rewriting Section 112 in Title III of the 1990 amendments. Congress defined 189 chemicals as hazardous (subject to possible deletion) and required technology-based controls on sources of those chemicals, as well as any others that might be added to the list by EPA. Sources that emit hazardous air pollutants will be regulated in two stages. In the first, technology-based emissions standards will be imposed. Each major source (defined, generally, as a stationary source having the potential to emit 10 tons per year of a single hazardous air pollutant or 25 tons per year of a combination of hazardous air pollutants) must meet an emission standard based on using the maximum available control technology (MACT) as defined by standards to be issued by EPA. Smaller sources, known as area sources, must meet emissions standards based on using generally available control technology.

Section 112 defines some contexts in which quantitative risk assessment will remain important. First, quantitative risk assessment will be relevant in determining which categories of sources will not be subject to technology-based regulation; EPA may delete a source category from regulation if no source in the category poses a risk of greater than 10^{-6} to the "individual most exposed to emissions." Even here, judging from the use of the word "may," EPA is not required to make the deletion; thus, the results of the quantitative risk assessment need not be decisive.

Quantitative risk assessment has a greater, but still limited, role in the second stage of standard-setting under Section 112(f), the "residual-risk" stage. That section requires EPA to set standards that protect public health with an ample margin of safety if it concludes that the first stage of technology-based standard-setting has not done so. Second-stage standards must be set for a category of "major sources" if the first stage allows a residual risk of greater than 10^{-6} to the individual most exposed to emissions. This requirement might seem a wholesale adoption of risk management based on the maximally exposed person, but two points must be noted. First, the 10^{-6} criterion for standard-setting need only be an upper-limit screening device. EPA is free, if it chooses, to set second-stage standards for source categories posing lesser risks. Second, the actual second-stage standard need not be expressed in terms of quantitative risk. Section 112(f)(2) authorizes EPA to continue the $10^{-4}/10^{-6}$ approach described earlier, but it does not require the agency to do so. Instead, any method is acceptable that comports with *NRDC v. EPA*'s requirement that the standards provide an "ample margin of safety" in addition to reducing risk to a level judged acceptable by EPA.

Such techniques as hazard assessment, hazard ranking, and risk ranking (discussed above), and in some cases quantitative risk assessment, can also play a role in the agency's decisions on questions such as these:

- *Should EPA modify the definition of "major source" to include sources emitting less than the statutory cutoffs?* Section 112(a) defines a major source as one with the potential to emit 10 tons per year of any single listed hazardous air pollutant or 25 tons of any combination of listed pollutants, but allows EPA to lower these thresholds for a pollutant on the basis of such factors as potency, persistence, and potential for bioaccumulation.
- *Should EPA list additional pollutants as hazardous or remove some pollutants from the list?* Section 112(b) establishes a list of 189 hazardous air pollutants and requires that EPA add a substance to the list on a determination, either on its own accord or in response to a petition, that the substance is "known to cause or may reasonably be anticipated to cause adverse effects to human health or adverse environmental effects." This standard represents a reaffirmation of the *Ethyl* decision (discussed later) that EPA may regulate in the face of scientific uncertainty about a substance's effects. EPA is required to delete a substance if it decides that data are adequate to show that the substance will not cause, or be reasonably anticipated to cause, an adverse effect. In deletions as well, the risks of uncertainty are put on the source.
- *Which sources of hazardous air pollutants ought EPA to regulate first?* Section 112 requires that EPA set technology-based standards for categories of major sources on a phased schedule beginning in 1992 and ending in 2000. In deciding the order in which standards will be set, EPA must consider known or expected adverse effects of the pollutants to be regulated, as well as the quantity and location of emissions, or reasonably anticipated emissions, of hazardous air pollutants in each category. EPA has completed this preliminary task (see EPA, 1992a).
- *What restrictions ought EPA to place on offsetting within plants?* Generally, a physical change at a plant that increases emissions of a hazardous air pollutant will subject the plant to special new-source requirements. Under Section 112(g), this will not be the case if the plant simultaneously decreases by an offsetting amount emissions of a more hazardous pollutant. Deciding which offsets, if any, qualify for Section 112(g) may require EPA to rank the relative potency of hazardous air pollutants.
- *What restrictions ought EPA to place on offsetting by sources seeking to qualify for the early-reduction program?* The "early-reduction" program will pose similar issues. Usually, a source will have up to 3 years to comply with an EPA standard for controlling hazardous air pollutants. A source can obtain a 6-year extension, however, if it shows that it has achieved by approximately the end of 1993 a reduction of at least 90% in emissions of hazardous air pollutants (95% for particulate hazardous air pollutants) from baseline emissions. EPA is required to disqualify reductions that were used to offset increases in emissions of pollutants for which high risks of adverse health effects might be associated with exposure even to small quantities. Here, too, EPA will have to grapple with the relative potency factors of hazardous air pollutants. These rules have already been issued (see EPA, 1992b).
- *Which substances should EPA attempt to control through its urban-area source program?* EPA is required to identify at least 30 hazardous air pollutants that, as the result of emissions from area sources (nonmajor sources other than vehicles or off-road engines), present the greatest threat to public health in the largest number of urban areas. The agency must also

identify categories responsible for those emissions and develop a national strategy that accounts for over 90% of the emissions of the identified air pollutants and that reduces by at least 75% the incidence of cancer attributed to exposure to hazardous air pollutants emitted by major and area sources.

- *Which pollutants ought EPA control under its authority to protect against accidental releases?* EPA must promulgate a list of 100 substances that, in the event of accidental release, are known to cause or can reasonably be expected to cause death, injury, or serious adverse effects to human health or the environment. The agency must also establish a "threshold quantity" for each. Operators of sources at which a listed substance is present in more than a threshold quantity must prepare a risk-management plan to prevent accidental releases.

Noncancer Risk Associated With Hazardous Air Pollutants

The current EPA approach to risk assessment for noncancer hazards posed by hazardous air pollutants, refined in several ways, is conceptually similar to the traditional approach to threshold agents described earlier. The agency identifies a so-called inhalation reference concentration (RfC). An RfC is defined by EPA as "an estimate (with uncertainty) of the concentration that is likely to be without appreciable risk of deleterious effects to the exposed population after continuous, lifetime exposure" (EPA, 1992b). RfCs are derived from chemical-specific toxicity data. The latter are used to identify the most sensitive indicator of a chemical's toxicity and the so-called no-observed-adverse-effect level (NOAEL) for that indicator effect. If the NOAEL is derived from an animal study, as is typically the case, it can be converted to a human equivalent concentration by taking into account species differences in respiratory physiology. Uncertainty factors, whose magnitudes depend on the nature of the toxic effect and the quantity and quality of the data on which the NOAEL is based, are applied to the human-equivalent NOAEL to estimate the RfC. That procedure is used for all forms of toxic hazard except carcinogenicity. The use of RfCs depends on the assumption that toxic effects will not occur until a threshold dose is exceeded (EPA, 1992b).

Another important provision of Title III of the 1990 Amendments was the requirement that environmental effects be included in the evaluation of a risk associated with a pollutant. An adverse environmental effect is defined in Section 112(a)(7) of the act as "any significant and widespread adverse effect, which may reasonably be anticipated, to wildlife, aquatic life, or other natural resources, including adverse impacts on populations of endangered species or significant degradation of environmental quality over broad areas." Appendix III of EPA's *Unfinished Business* report (EPA, 1987b) found that airborne toxic substances, toxic substances in surface waters, and pesticides and herbicides were in the second highest category of relative risk in the ecological and welfare categories. Of particular concern in this report was the transport by air and water of toxic substances (heavy metals and organics) that accumulate

in ecological food chains. Such bioaccumulation has impacts on both ecological resources and the use by humans of specific ecological populations (e.g., fish consumption). Ecological risk assessment is not discussed in this report except to the extent that bioaccumulation affects the health of people who eat and drink contaminated ecological resources, but is discussed in another recent NRC report entitled *Issues in Risk Assessment* (NRC, 1993a).

PUBLIC CRITICISM OF CONDUCT AND USES OF RISK ASSESSMENT

The development of risk-assessment methods and their expanding uses in the federal and state regulation of hazardous substances have been carefully scrutinized by interested parties in the regulated industries, environmental organizations, and academic institutions. That scrutiny has led to frequent and sharp criticisms of the methods used for assessing risk and of ways in which the results of risk assessment have been used to guide decision-making. The criticisms have not been directed solely at the use of risk assessment in regulation of hazardous air pollutants, but rather cover a range of uses.

We cite here some of the criticisms that have appeared in the literature or that have otherwise been presented to the committee, because they help to define the issues reviewed in this report. *We emphasize that our citation of these criticisms does not mean that we believe them to be valid. Nor is the order of their listing meant to suggest our opinion regarding their possible importance.*

CRITICISMS PERTAINING TO CONDUCT OF RISK ASSESSMENT

(1) Some analysts have commented that the default options used by EPA (i.e., the science-policy components of risk assessment) are excessively "conservative" or are not consistent with current scientific knowledge. The cumulative and combined effect of the many conservative default options used by EPA might yield results that seriously overstate actual risks, and thus tend to overcontrol emissions.

(2) Some experts have noted that important aspects of risk are neglected by EPA. The agency does not appear to recognize the possibility of synergistic interactions when multiple chemical exposures occur, nor does it seem concerned that available data show extreme variability among individuals in their responses to toxic substances. The failure to deal with those issues can lead to serious underestimation of human risk, especially at very low exposures. A related issue is the overlooked problem of risk aggregation—how risks associated with multiple chemicals are to be combined.

(3) The default options used by EPA have, according to some, become excessively rigid.

The barriers to using alternative assumptions by incorporating chemical-specific data are said to be in effect impassable, because the degree of scientific certainty has never been explicitly or implicitly defined by EPA. The too-rigid adherence to the preselected default options also impedes research, because there is little likelihood that novel data will be incorporated into EPA risk assessments.

(4) Many commentators have stated that insufficient attention has been paid to the issue of human exposure itself. In particular, EPA has not defined the terms of exposure assessment with sufficient clarity. How are populations and subpopulations of interest to be characterized? What is meant by such terms as "maximally exposed individual" and "reasonable maximal exposure"? How are multiple exposure pathways to be assessed in evaluating individual's total risk associated with a hazardous air pollutant?

(5) Some have noted that the uncertainties in the results of risk assessments are inadequately described. Risks are most often reported as "point estimates," single numbers that admit to no uncertainty. Large uncertainties are often overlooked, and descriptions of risk as "upper bounds" are misleading and simplistic.

(6) According to some, insufficient attention has been devoted to noncancer risks. The NOEL-safety factor approach, although useful, is not scientifically rigorous.

(7) Some believe that we do not have sufficient knowledge to make risk estimates. In addition, some believe that a risk assessor can make risk calculations come out high or low, depending on what answer is desired. Thus, some people believe that credible risk assessment might be impossible to obtain with the existing state of science and risk-assessment institutions.

Criticisms Pertaining to the Relationship Between Risk Assessment and Risk Management

(1) Several commentators have concluded that the conceptual separation of risk assessment and risk management called for in the 1983 NRC report has resulted in procedural separation to the detriment of the process. Some commentators have viewed the publication of toxicity values (cancer potency factors and reference doses) by one office of EPA for the use of other offices (those responsible for regulatory decision-making) as a prime example of undesirable separation.

(2) According to some analysts, upper-bound point estimates of risk, produced solely for screening or risk-ranking purposes, have too often been used inappropriately as a definitive basis for decision-making. Such use might be attractive to decision-makers, but it seriously distorts the intentions of risk assessors who produce the estimates. Managers need to consider scientific uncertainties more fully.

(3) Several commentators have expressed the view that risk assessment is too resource-intensive and thus impedes action. Given the substantial uncertainties in the results of risk assessment, it seems inappropriate to devote so much effort to its conduct. Moreover, no good mechanisms exist to resolve controversies, so debates over the appropriateness of various risk-assessment outcomes can be endless.

(4) Some reviewers, particularly those with state governments, believe that more effort needs to be devoted to defining the uses to which a risk assessment is to be put before it is attempted. Such planning will help to deal with the problem of resource allocation, because the amount of effort needed for a risk assessment can be more appropriately matched to its ultimate uses.

(5) Some analysts have pointed out that the failure to pay sufficient attention to the results of risk assessment has resulted in misplaced priorities and regulatory actions that are driven by social forces, not by science. They note that the fact that risk assessment is imperfect does not justify the use of decision-making approaches that suffer from even greater imperfections.

(6) On the other hand, some commentators feel that risk assessment has been given too much weight, especially in light of its methodological limitations and inability to account for unquantifiable features of risk, such as voluntariness and fear.

(7) Some analysts also point out that far too little attention has been devoted to research to improve risk-assessment methods. It is unfair simply to criticize the methods without offering risk assessors the means to improve them.

Are any of those criticisms justified? If so, what responses can be made to them? Can improvements be made? If so, how will they affect the conduct of risk assessment and the use of risk-assessment results in regulatory decision-making? These and related issues are the primary focus of Chapters 6-12 of this report.

3
EXPOSURE ASSESSMENT

INTRODUCTION

Accurate information on human exposure to hazardous air pollutants emitted by various sources is crucial to assessing their potential health risks. This chapter describes methods used to assess exposure to hazardous air pollutants. Section 112 of the Clean Air Act Amendments of 1990 applies to major sources that either singly or in combination emit defined quantities of one or more of the 189 hazardous air pollutants. The sources to which the act applies emit pollutants both continuously and episodically, and the pollutants can move from air to water, soil, or food.

In the terminology of the Environmental Protection Agency (EPA) and Title III of the 1990 Amendments, a major source of pollution is considered to be

> any stationary source or group of stationary sources located within a contiguous area and under common control that emits or has the potential to emit considering controls, in the aggregate, 10 tons per year or more of any hazardous air pollutant or 25 tons per year of any combination of hazardous air pollutants. The [EPA] Administrator may establish a lesser quantity, or in the case of radionuclides different criteria, for a major source than that specified in the previous sentence, on the basis of the potency of the air pollutant, persistence, potential for bioaccumulation, other characteristics of the air pollutant, or other relevant factors.

A stationary source is "any building, structure, emission source, or installation which emits or may emit any air pollutant."

As part of determining the health threat of a pollution source to humans, EPA assesses how a pollutant moves from a source through the environment until it makes contact with humans in its original form or after conversion to other substances. For most airborne substances, inhalation is assumed to be the primary route of entry into the body. There has recently been an extensive review of advances in assessing human exposure to airborne constituents (NRC, 1991a). That review attempted to define exposure carefully as a part of the overall continuum that leads to illness brought about by environmental contaminants. The definition of exposure as a part of this continuum has been incorporated into the 1992 revised guidelines for exposure assessment developed by EPA (1992a).

Human exposure to a contaminant is an event consisting of contact with a specific contaminant concentration at a boundary between a human and the environment (e.g., skin or lung) for a specified interval; total exposure is determined by the product of concentration and time.

The amount of a substance that is absorbed or deposited in the body of an exposed person in a given period is the administered dose. Calculating the dose from the exposure depends on a number of factors, including the mode of entry into the body. For substances that move into the body through an opening—such as the mouth or nose via breathing, eating, or drinking—the dose depends on the amount of the carrier medium that enters the body. For airborne substances, the potential dose is the product of breathing rate (volume of air inhaled per unit of time), exposure concentration, and fractional deposition of the substance throughout the respiratory tract. However, an inhalation exposure will not lead to a dose if none of the substance is absorbed through the lung or deposited on the surface of the lung or other sections of the respiratory tract.

A pollutant can also enter the body through the skin or other exposed tissues, such as the eyes. The substance is then directly absorbed from the carrier medium into the tissue, often at a rate that is different from the rate of absorption of the carrier. The pollutant uptake rate is the amount of the pollutant absorbed per unit of time, and the dose is the product of exposure concentration and uptake rate at that concentration. The NRC report on exposure assessment (NRC, 1991a) provides a scientific framework to identify routes of entry and degree of contact and indicates how exposure assessment integrates data on emitted pollutants with biological effects.

Exposure assessment involves numerous techniques to identify a pollutant, pollutant sources, environmental media of exposure, transport through each medium, chemical and physical transformations, routes of entry to the body, intensity and frequency of contact, and spatial and temporal concentration patterns of the pollutant. Mathematical models that can be used to describe the relationships among emissions, exposures, and doses are shown in Appendix C.

Exposure to a contaminant can be estimated in three ways. It can be evaluated directly by having a person wear a device that measures the concentration of a pollutant when it comes into contact with the body. Environmental monitoring is an indirect method of determining exposure, in which a chemical's concentration is measured in an environmental medium at a particular site, and the extent to which a person is exposed to that medium is used to estimate exposure. Finally, exposure can be estimated from the chemical's actual dose to the body, if it manifests itself in some known way through a measurable internal indicator (biological marker), such as the concentration of the substance or its metabolite in a body tissue or excreted material (NRC, 1991a). This is a direct method of exposure estimation and, unlike the other two, accounts for the amount of contaminant absorbed by the body. Each of these methods provides an independent estimate of exposure; when it is possible to use more than one approach, comparison of results can be useful in validating exposure estimates.

EPA's air-pollution regulatory programs have relied primarily on mathematical models to predict the dispersion of emissions to air and the potential for human exposure under different emission-control scenarios (see Appendix C for a description of EPA's Human Exposure Model). Source-emission estimates and meteorologic data were used to calculate the expected long-term ambient concentrations at various distances and directions from the source. Census data were used to estimate the number and location of people living near the source. A high-exposure scenario was estimated for a person (e.g., maximally exposed individual, MEI)

assumed to be living near the source and constantly exposed for 70 years to the highest estimated air-pollutant concentration. EPA does not modify exposure estimates by including mobility of the population, shielding due to indoor locations, or additional exposures from indoor or other community sources. EPA also used a modeling approach to estimate the exposure of the local population to an average concentration of pollutant emitted from a source (EPA, 1985a).

1992 Exposure-Assessment Guidelines

EPA has recently promulgated a new set of exposure-assessment guidelines to replace the previous (1986) version (EPA, 1992a). The approach in the new guidelines is very different from that in the previous version and generally follows many of the concepts of exposure assessment presented in the 1991 NRC report (NRC, 1991a). The guidelines explicitly consider the need to estimate the distribution of exposures of individuals and populations and discuss the need to incorporate uncertainty analysis into exposure assessment. This approach is consistent with the most recent NRC recommendations on exposure analysis (NRC, 1993e).

The guidelines discuss the roles of both analytic measurement and mathematical modeling in estimating concentrations and durations of exposure. They do not recommend specific models, but suggest that models match the objectives of the particular exposure assessment being conducted and that they have the accuracy needed to achieve those objectives. They also call for detailed explication of the choices and assumptions that often must be made in the face of incomplete data and insufficient resources.

Exposure Calculation and the Maximally Exposed Individual

EPA has traditionally characterized exposure according to two criteria: exposure of the total population and exposure of a specified, usually highly or maximally exposed individual. The MEI's exposure is estimated as the plausible upper bound of the distribution of individual exposures. The reason for finding the MEI, as well as population exposure, is to assess whether any individual exposure might occur above a particular threshold that, as a policy matter, is considered to be important. Because the MEI's exposure level is intended to represent a potential upper bound, its calculation has involved a variety of conservative assumptions. Among the more conservative, and more contentious, were that the MEI lived for 70 years at the location deemed by the dispersion model to receive the heaviest annual average concentration, that the person stayed there 24 hours/day, and that there is no difference between outdoor and indoor concentrations. In practice, it is straightforward to estimate the exposure of an immobile MEI with the air-quality models described below. However, estimating exposure for a more typical person requires much more information as to his or her activities during the assessment period. Usually, these activities include spending a majority

of time inside (where pollutant concentrations can be attenuated) and time spent in travel away from the residence. The 70-year, 24-hour/day and no-indoor-attenuation assumptions are, in effect, bounding estimates. Some people do live in a small community for a whole lifetime. Some people do spend virtually their whole life at home. And for some pollutants, there is little attenuation of pollutant concentrations indoors. Nonetheless, the occurrence of these conditions is rare, and it is even rarer that all these are found together.

In the most recent exposure guidelines, EPA no longer uses the term MEI, noting the difficulty in estimating it and the variety of its uses. The MEI has been replaced with two other estimators of the upper end of the individual exposure distribution, a "high-end exposure estimate" (HEEE) and the theoretical upper-bounding estimate (TUBE). The HEEE is not specifically defined ("the Agency has not set policy on this matter" [EPA, 1992a]); rather, the new exposure guidelines discuss some of the issues and procedures that should be considered as part of the choice of the methods and criteria. The HEEE is "a plausible estimate of exposure of the individual exposure of those persons at the upper end of an exposure distribution." *High end* is stated conceptually as "above the 90th percentile of the population distribution, but not higher than the individual in the population who has the highest exposure." As is implied by those statements, the new guidelines have adopted the use of individual exposure distributions, and the HEEE is a value in the upper tail of that distribution. The exact percentile for the HEEE that should be picked from the exposure distribution is not specified, but, according to EPA, should be chosen to be consistent with the population size in the particular application. The TUBE is a "bounding calculation that can easily be calculated and is designed to estimate exposure, dose, and risk levels that are expected to exceed the levels experienced by all individuals in the actual distribution. The TUBE is calculated by assuming limits for all the variables used to calculate exposure and dose that, when combined, will result in mathematically highest exposure or dose. . . ." In addition, calculation of the TUBE includes using a limiting case for the exposure-dose and dose-response relationships in calculating risk.

To be responsive to the concerns raised in the NRC (1991a) report, EPA changed its approach to the MEI. The TUBE is to be used only for bounding purposes and is to be superseded by the HEEE in detailed risk characterizations. Although the exposure guidelines are ambiguous in details about the determination of the HEEE, the HEEE is based on the estimation of the distribution of exposures that people might actually encounter. From the individual exposures, it is possible to develop population exposure (and risk) distributions and include uncertainty estimation, and personal-activity patterns. The details of these approaches are discussed in the applicable sections of this report (Chapters 10, 11, and 12).

The calculation of the exposure distribution for an individual requires knowledge of both the distribution of hazardous-pollutant concentrations and the distribution of times that the individual spends in places for which the concentrations are measured or modeled (time-activity patterns). For estimates of population exposure, the individual time-activity patterns are estimated for the population of the individuals that might be exposed.

EMISSION CHARACTERIZATION

The first step in exposure assessment is estimation of the quantity of toxic materials emitted by a given source. Emission characterization involves identifying the chemical components of emissions and determining the rates at which they are emitted. Although emission characterization is a necessary part of the exposure-assessment process, it is often conducted separately from exposure assessment to determine whether a given operation falls into one or another regulatory category.

SOURCES OF EMISSIONS

The emission rate often is considered to be proportional to the type and magnitude of industrial activity at a source. Emissions from a source might occur from process vents, handling equipment such as valves, pumps, etc., storage tanks, transfer, and wastewater collection and treatment. Process-vent emissions are released to the atmosphere from the use, consumption, reaction, and production of chemicals. Fugitive emissions are produced when chemicals "escape" from handling equipment, such as pumps and valves. Storage-tank emissions are released from the locations where chemical feedstocks or products are stored. These emissions depend on the chemical properties of the product stored (e.g., the vapor pressure), the atmospheric conditions (e.g., temperature), the type of tank (e.g., fixed or floating roof), and the type of seal and venting used. Transfer emissions are produced as material is received from or loaded into storage tanks, tank trucks, rail cars, and marine vessels (e.g., barges and ships). When material is added to a storage tank, for example, it can displace contaminated air into the atmosphere. Wastewater collection and treatment emissions can be released into a plant's wastewater system when chemicals are processed and released from the wastewater treatment plant. In continuous processes, a malfunction (upset), startup, or shutdown of the process can result in a much greater emission than normal.

EMISSION ESTIMATION METHODS

EPA (1991c) has provided a detailed procedure for estimating the emissions from facilities that use hazardous chemicals. In estimating emissions, information is generally needed on the magnitude of use of given chemicals, the chemical characteristics of the chemicals, and the efficiency with which the emissions are controlled.

The EPA protocols (1991c) provide a tiered approach to emission estimations ranging from relatively simple emission factors to material balances and direct measurements. These approaches have varied accuracy in estimation and a wide range of costs.

An emission factor is a multiplication factor that allows determination of the average emissions likely to come from a facility on the basis of its level of activity (EPA, 1985b). Emission factors are calculated on the basis of average measured emissions at several facilities

in a given industry (*Compilation of Air Pollutant Emission Factors,* commonly known as AP-42 [EPA, 1985b]).

A material balance is performed by assuming that the sum of the mass of chemical inputs minus the sum of the outputs, after all chemical changes and accumulation within the process or equipment have been accounted for, is the emission. In general, material balances produce information about emissions that depends on relatively small differences between the large numbers that characterize inputs (raw materials) and outputs (finished products, byproducts, and other wastes).

Emissions can be estimated with calculation methods presented in EPA (1988d) publications, such as *Protocols for Generating Unit-Specific Emission Estimates for Equipment Leaks of VOC and VHAP* (used for fugitive emissions). This emission-estimation method allows the development of site-specific emission factors based on testing a statistical number of sources at a facility. These site-specific emission factors can be used to develop emission estimates in the future.

Ideally, emissions from a source can be calculated on the basis of measured concentrations of the pollutant in the source and the emission rate of the source. This approach can be very expensive and is not often used. The emission rates, characteristics of the source facility (stack height, plume temperature, etc.), and local topography (flat or complex terrain) are used to estimate the ambient concentrations of the hazardous pollutants to which people can be exposed.

MEASUREMENT METHODS

The concentration of a given pollutant can be measured in each microenvironment. A microenvironment is a three-dimensional space with defined boundaries of which contaminant concentration is approximately spatially uniform during some specific period (Sexton and Ryan, 1988). There have been substantial improvements in analytic methods to measure concentrations, as described in a 1991 NRC report (NRC, 1991a). Modern methods in computerization of instruments, data recording, and data processing also permit much greater capability to obtain detailed information on the temporal and spatial variability of contaminants over a range of microenvironments. Other substantial improvements have enhanced the utility of personal monitors, which are worn by subjects directly and record the concentration or collect time-integrated samples of specific pollutants with which the wearers come into contact for specific intervals. For example, assessment of exposure to radiation has long made use of inexpensive, accurate, integrating dosimeters that were first developed when research on radioactive materials and the use of radioactivity were expanding rapidly. There are often substantial variations in the spatial distribution of radiation within a microenvironment, so individual dosimeters have been thought to provide the best estimates of individual exposure. Individual monitoring and extensive microenvironmental measurements are not generally practical for assessing exposures of the general population, but because of cost and the unwillingness of individuals to participate in exposure assessments, new instruments, including

passive dosimeters for airborne chemicals, are likely to permit a similar strategy. These methods have been used in the TEAM studies (Wallace, 1987) to examine the total exposure of individuals to a number of volatile organic compounds in several locations around the country. This approach to exposure assessment has been applied in other research studies. One important finding of the TEAM studies (and others) is that substantially greater exposures to many contaminants occur indoors, both because of the higher concentrations and because most people spend considerably more time inside.

Although field measurement studies are generally expensive and require careful planning, organization, and quality-assurance programs, measurement programs can provide the large amounts of high-quality data needed to characterize environmental systems, to estimate exposure, and to develop, test, and evaluate models for evaluating exposure. Documented reliable models can then be used in place of more expensive, direct measurements. Reliable measurements are generally needed to provide knowledge of emissions of chemicals that give rise to human exposures. However, measurements provide only information on the current status of the system. To allow for a broader range of meteorologic conditions, estimate the effects of changes in plant operating capabilities and procedures, or estimate the effects of an accident or upset condition, models are needed to estimate emissions and the transport of emitted materials in the atmosphere.

MODELING USED IN EXPOSURE ASSESSMENT

Mathematical models used in exposure assessment can be classified in two broad categories: models that predict exposure (in units of concentration multiplied by time) and models that predict concentration (in units of mass per volume). Exposure models can be used to estimate population exposures from small numbers of representative measurements. Although concentration (or air-quality) models are not truly exposure models, they can be combined with information on human time-activity patterns to estimate exposures.

Air-quality models are also used to predict the fate, such as deposition or chemical transformation, of atmospheric pollutants to which people can be exposed indirectly (e.g., through deposition of pollutants from air onto surface water followed by bioaccumulation in fish). Such models are central to risk assessment (see Figure 3-1). They constitute the only method of determining the total impact of diverse emissions on air quality and are key tools in assessing the impact of specific sources on future air-pollutant concentrations and deposition.

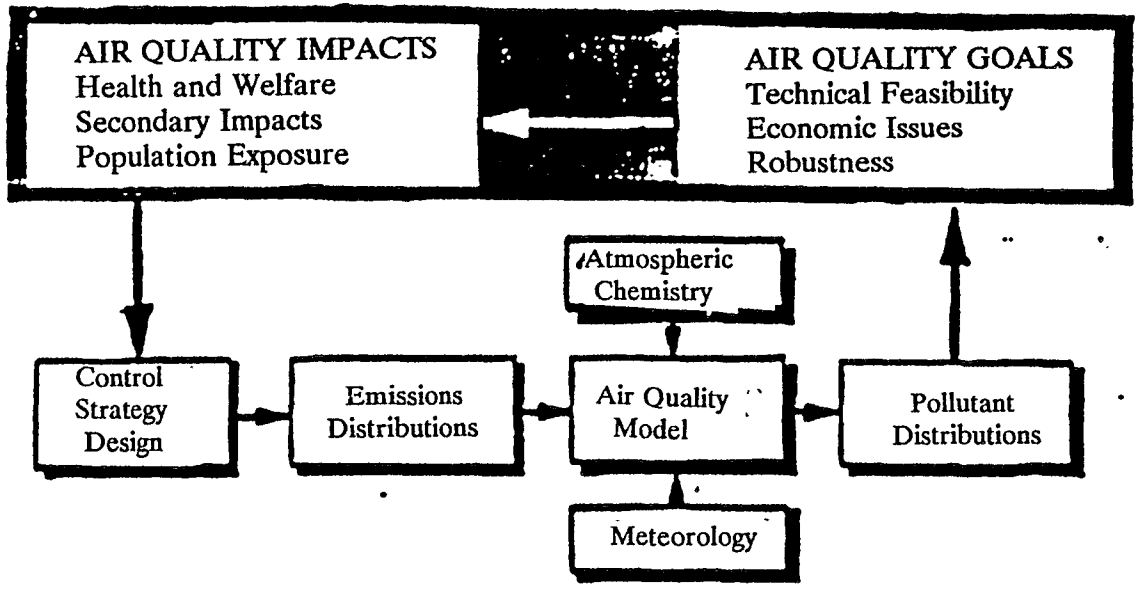


Figure 3-1 Air quality control strategy design process.

Modeling Airborne Concentrations

Mathematical air-quality models used in air-pollution analysis are in two classes: empirical and analytic. The former type of model statistically relates observed air quality to the accompanying emission patterns, with chemistry and meteorology included only implicitly. Although they hold promise for use in some aspects of air-pollutant risk assessment, these models are not commonly used by EPA in its risk-assessment practice and will be discussed later. EPA and others more commonly use the form of analytical models, in which analytic or numerical expressions describe the complex transport processes and chemical reactions that affect air-pollutant concentrations. Pollutant concentrations are determined as explicit functions of meteorologic and topographic characteristics, chemical transformation, surface deposition, and source characteristics. In exposure assessments of air pollutants, the most widely used set of models has been the class called Gaussian-plume models. Gaussian-plume models are derived from atmospheric diffusion theory assuming stationary, homogeneous turbulence or, alternatively, by solution of the atmospheric-diffusion equation assuming simplified forms of the effective diffusivity (Seinfeld, 1986). Within the limits of the simplifications involved in their derivation, they can describe the individual processes that affect pollutant concentrations, such as diffusion, bulk transport by the wind, and deposition. These models are a type of a much broader family of models called dispersion or atmospheric-transport models. See Appendix C for more information.

Modeling Multimedia Exposure to Air Pollutants

In some cases, exposure to toxic pollutants emitted into the atmosphere occurs by pathways other than, or in addition to, inhalation. An example is deposition of metals like mercury in surface waters followed by the bioaccumulation of methyl mercury in fish and then ingestion of contaminated fish. Another is exposure of an infant ingesting the breast milk of a mother exposed to a toxic pollutant, such as polychlorinated biphenyls; this can be an important route for lipophilic compounds (NRC 1993e), and EPA has investigated it in some exposure assessments. Recent studies (Travis and Hattemer-Frey, 1988; Bacci et al., 1990; Trapp et al., 1990) have also found significant bioaccumulation of chemicals from the atmosphere in plant tissues, particularly of nonionic organic compounds. These studies have found that the degree of bioaccumulation depends on solubility, and models for the uptake have been developed (Stevens, 1991). Such "indirect" pathways can concentrate pollutants and thus result in significant increases in exposure.

Multimedia exposure and indirect exposure have been considered more frequently in hazardous-waste site (e.g., Superfund) cleanup than in the management of exposure to industrial air pollutants. One example of multiple-path exposure to a source of primary air pollutants conducted by EPA is found in Cleverly et al. (1992). Multiple air pollutants, including heavy metals and organic chemicals, were followed after emission from a municipal-waste combustor. Atmospheric transport and deposition were modeled with a Gaussian-plume

model modified to include wet and dry deposition. Other models were used to assess pollutant concentrations in nearby bodies of water; bioaccumulation; consumption of animal tissue, plants, and water; soil ingestion; and total potential dose.

Alternative Transport and Fate Models

The 1992 EPA guidelines for exposure assessment offer an approach to selection and use of models to estimate transport and fate, as well as exposure, so a variety of models can be used. For rapid screening analyses, Gaussian-plume models are adequate for limited distances around the source. However, for a more complete characterization of the distribution of concentrations downwind of a source, more refined modeling approaches may be needed.

In recent years, stochastic modeling of atmospheric dispersion has increased in popularity because of its relatively simple concept, its applicability to more complicated problems, and the improvements in computer capability and costs that make such models practical. Stochastic models can easily incorporate real physical phenomena, such as buoyancy, droplet evaporation, variations in the dispersity of released particles, and dry deposition. Stochastic modeling is typically implemented as a numerical Monte Carlo model in which the movement of a large number of air parcels is tracked in a Lagrangian reference frame. The concentration profile is then obtained from the air-parcel positions.

Boughton et al. (1987) described a Monte Carlo simulation of atmospheric dispersion based on treating either parcel displacement or parcel velocity as a continuous-time Markov process (a one-step-memory random process like Brownian movement). They simplified the problem by restricting themselves to crosswind-integrated point sources and assumed that dispersion in the mean wind direction is negligible. Thus, they reduced the problem to a one-dimensional model. Liljegren (1989) extended the model to incorporate both horizontal and vertical dispersion perpendicular to the mean wind direction. He found good agreement between the results of the three-dimensional stochastic model with concentration data found in the literature. Recent measurements of the dispersion of ground-released smokes and obscurants have shown excellent agreement of his stochastic model both with the average concentration values, including the profile across the plume, and with the time-varying concentrations observed (pers. comm., W. E. Dunn, U. of Illinois, 1988). It appears from those results that stochastic models offer considerable improvement over conventional Gaussian-plume models. Thus, there will soon be a substantially improved ability to predict average and time-varying ground-level concentrations.

Time-Activity Patterns

Exposure occurs when someone is in contact with a substance for some period. To estimate exposures, it is necessary to estimate the time spent in various activities that provide the opportunity for exposure. Figure 3-2 shows one such analysis. Various methods are available (NRC, 1991a), including recording of activities in a time-use diary (which might be automated to facilitate the recording of locations at specific times of the day and might use questionnaires to help reconstructing kinds and duration of activities). Some participants are careful in recording their activities; others might not provide accurate accounts, because of oversight or carelessness. The framing and wording of questionnaires can substantially affect the results of a survey and thus bias the resulting estimates of time spent in various activities and locations. Further work in the measurement and modeling of time and activity is needed; research recommendations were presented in an earlier report (NRC, 1991a).

Exposure-Assessment Models

The 1992 guidelines call for the development of distributions, instead of point estimates, for exposure parameters. It is the exposure-prediction models that combine microenvironmental concentration estimates with information on time-activity patterns of people to estimate individual exposures or the distribution of individual exposures in a typical population. Activity patterns and microenvironmental concentrations can both be measured or modeled. Microenvironmental concentrations and activity patterns can vary from person to person, and from period to period. Three types of models have been developed to estimate population exposures: simulation models, such as the simulation of human air pollution exposure (SHAPE) model (Ott, 1981, 1984; Ott et al., 1988) and National Ambient Air Quality Standards (NAAQS) Exposure Model (NEM) (Johnson and Paul, 1981, 1983, 1984), the convolution model of Duan (1981, 1987), and the variance-components model of Duan (1988) and Switzer (1988) (see Appendix C for additional information). The development of total-exposure models is one of the advances in modeling.

Several of the models for predicting exposures assume some correlation between measured contaminant concentrations in a microenvironment and the time spent by the exposed person in that space. Studies by Duan et al. (1985) suggested, on the basis of data from the Washington, D.C., carbon monoxide (CO) study (Akland et al., 1985), that there is no correlation between CO concentrations and time. However, there will be problems in existing models if occupancy times and concentrations of other contaminants correlate, as they might for irritating toxicants, such as formaldehyde.

Current exposure models use a variety of crude assumptions about the constancy of concentrations in microenvironments, the human activity patterns that determine the amount of time people spend in each microenvironment, and how representative the sampled population is to the total population that might be exposed to a contaminant.

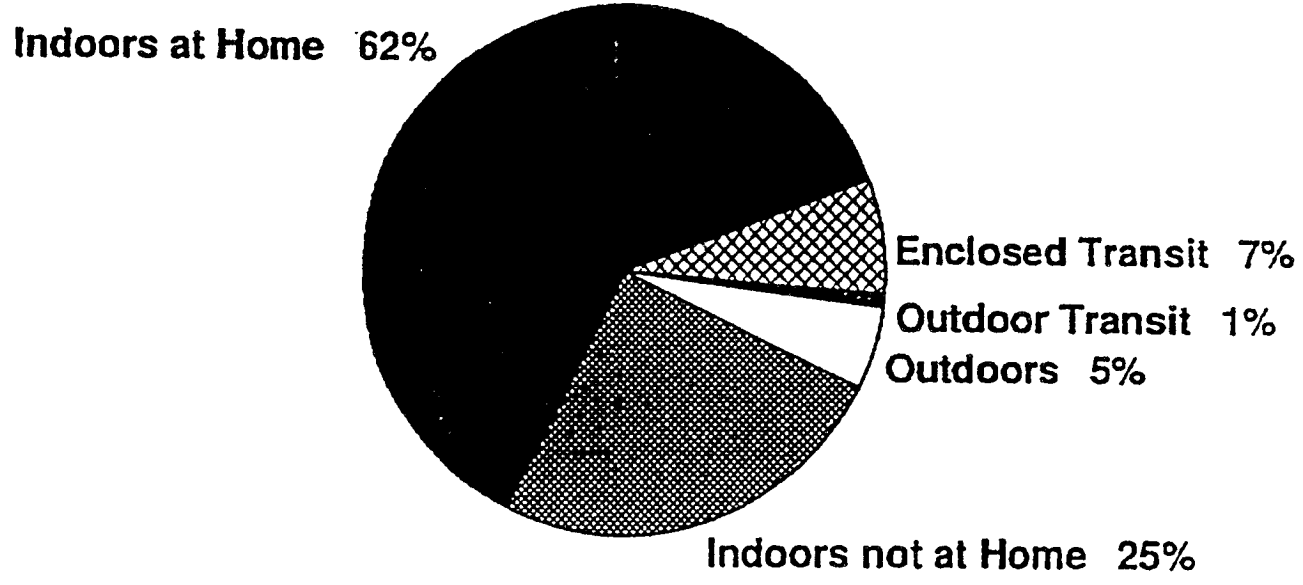


Figure 3-2 Percentage of day spent in different locations. Californians > 11 years of age. (Population Means). Source: Reprinted from Atmos. Environ. 26A, P. Jenkins, T.J. Philips, E.I. Mulberg, and S.P. Hui, Activity patterns of Californians, pp. 2141-2148, 1992.

Long-Term Exposure Modeling

Modeling very-long-term exposures, as is required for cancer risk assessment, presents several major difficulties. The current practice is to measure or model the concentration of a contaminant at one time and determine lifetime exposure by multiplying that concentration by a fixed number of years, e.g., the lifetime of an exposed person. However, the nature of exposure sources (e.g., changes in industrial processes) and activity patterns can change substantially over a lifetime. New sources or uses of sources can be introduced into the environment (e.g., the spreading use of wood-burning stoves), and old sources can be eliminated or modified (e.g., by the use of catalytic converters in motor vehicles). Typically, large facilities have a design life of 30 years, so considerable change can be anticipated in sources over the 70 years of a typical lifetime-exposure calculation.

Time-activity patterns of people can also vary substantially over very long periods. In the United States, people generally change their place of residence frequently, although some live in the same place over a lifetime. Population mobility can have a large impact on exposure assessments of agents, such as radon, that require reasonable estimates of long-term and highly variable exposure concentrations.

A person's activity pattern changes from childhood through young adulthood to middle and old age. Some efforts have addressed age-related differences in exposure that arise because of age. However, that aspect of variability in exposure over long periods has generally not received much attention in exposure modeling.

Short-Term Exposure Modeling

The typical steady-state airborne-concentration models are not able to provide estimates below 1-hour averages and have difficulty in modeling concentrations that vary widely over time and that can lead to short-term high exposures. If an exposure model is to estimate the effects of peak exposures on sensitive populations, the concentration model must provide reliable estimates for the time scales needed. There have been some important developments in stochastic models that could provide such estimates, but these developments have not yet been incorporated into the procedures for estimating exposure.

4

ASSESSMENT OF TOXICITY

INTRODUCTION

This chapter discusses the methods used to evaluate the toxicity of a substance for the purpose of health risk assessment. Evaluation of toxicity involves two steps: hazard identification and dose-response evaluation. Hazard identification includes a description of the specific forms of toxicity (neurotoxicity, carcinogenicity, etc.) that can be caused by a chemical and an evaluation of the conditions under which these forms of toxicity might appear in exposed humans. Data used in hazard identification typically are derived from animal studies and other types of experimental work, but can also come from epidemiologic studies. Dose-response evaluation is a more complex examination of the conditions under which the toxic properties of a chemical might be evidenced in exposed people, with particular emphasis on the quantitative relationship between dose and toxic response. This step also includes study of how response can vary from one population subgroup to another.

PRINCIPLES OF TOXICITY ASSESSMENT

The basic principles guiding the assessment of a substance's toxicity are outlined in the *Guidelines for Carcinogen Risk Assessment* (EPA, 1987a) (currently being updated), *Chemical Carcinogens: A Review of the Science and Its Associated Principles* (OSTP, 1985), *Guidelines for Developmental Toxicity Risk Assessment* (EPA, 1991a) and have recently been summarized by the NRC (1993a). In addition, guidelines for the assessment of acute toxicity have recently been developed by NRC (1993b). The developmental-toxicity guidelines are used in this chapter to illustrate EPA's approach to health effects that involve noncancer end points. They constitute the first completed noncancer risk-assessment guidelines in a series that EPA plans to issue.

HAZARD IDENTIFICATION

The first of the two questions typically considered in the assessment of chemical toxicity concerns the types of toxic effects that the chemical can cause. Can it damage the liver, the

kidney, the lung, or the reproductive system? Can it cause birth defects, neurotoxic effects, or cancer? This type of *hazard* information is obtained principally through studies in groups of people who happen to be exposed to the chemical (epidemiologic studies) and through controlled laboratory experiments involving various animal species. Several other types of experimental data can also be used to assist in identifying the toxic hazards of a chemical.

EPIDEMIOLOGIC STUDIES

Epidemiologic studies clearly provide the most relevant kind of information for hazard identification, simply because they involve observations of human beings, not laboratory animals. That obvious and substantial advantage is offset to various degrees by the difficulties associated with obtaining and interpreting epidemiologic information. It is often not possible to identify appropriate populations for study or to obtain the necessary medical information on the health status of individuals in them. Information on the magnitude and duration of chemical exposure, especially that experienced in the distant past, is often available in only qualitative or semiquantitative form (e.g., the number of years worked at low, medium, and high exposure). Identifying other factors that might influence the health status of a population is often not possible. Epidemiologic studies are not controlled experiments. The investigator identifies an exposure situation and attempts to identify appropriate "control" groups (i.e., unexposed parallel populations), but the ease with which this can be accomplished is largely beyond the investigator's control. For those and several other reasons, it is difficult or impossible to identify cause-effect relationships clearly with epidemiologic methods (OSTP, 1985).

It is rare that convincing causal relationships are identified with a single study. Epidemiologists usually weigh the results from several studies, ideally involving different populations and investigative methods, to determine whether there is a consistent pattern of responses among them. Some of the other factors that are often considered are the strength of the statistical association between a particular disease and exposure to the suspect chemical; whether the risk of the disease increases with increasing exposure to the suspect agent; and the degree to which other possible causative factors can be ruled out. Epidemiologists attempt to reach consensus regarding causality by weighing the evidence. Needless to say, different experts will weigh such data differently, and consensus typically is not easily achieved (IARC, 1987).

In the case of chemicals suspected of causing cancer in humans, expert groups ("working groups") are regularly convened by the International Agency for Research on Cancer (IARC) to consider and evaluate epidemiologic evidence. These groups have published their conclusions regarding the "degrees" of strength of the evidence on specific chemicals (sometimes chemical mixtures or even industrial processes when individual causative agents cannot be identified). The highest degree of evidence—sufficient evidence of carcinogenicity—is applied only when a working group agrees that the total body of evidence is convincing with respect to the issue of a cause-effect relationship.

No similar consensus-building procedure has been established regarding other forms of toxicity. Some epidemiologists disagree with IARC's cancer classification judgments in particular cases, and there seems to be even greater potential for scientific controversy regarding the strength of the epidemiologic evidence of non-cancer (e.g., reproductive, developmental, etc.) effects. There has been much less epidemiologic study of other toxic effects, in part because of lack of adequate medical documentation.

ANIMAL STUDIES

When epidemiologic studies are not available or not suitable, risk assessment may be based on studies of laboratory animals. One advantage of animal studies is that they can be controlled, so establishing causation (assuming that the experiments are well conducted) is not in general difficult. Another advantage is that animals can be used to collect toxicity information on chemicals before their marketing, whereas epidemiologic data can be collected only after human exposure. Indeed, laws in many countries require that some classes of chemicals (e.g., pesticides, food additives, and drugs) be subjected to toxicity testing in animals before marketing. Other advantages of animal tests include the facts that

- The quantitative relationship between exposure (or dose) and extent of toxic response can be established.
- The animals and animal tissues can be thoroughly examined by toxicologists and pathologists, so the full range of toxic effects produced by a chemical can be identified.
- The exposure duration and routes can be designed to match those experienced by the human population of concern.

But laboratory animals are not human beings, and this obvious fact is one clear disadvantage of animal studies. Another is the relatively high cost of animal studies containing enough animals to detect an effect of interest. Thus, interpreting observations of toxicity in laboratory animals as generally applicable to humans usually requires two acts of extrapolation: interspecies extrapolation and extrapolation from high test doses to lower environmental doses. There are reasons based on both biologic principles and empirical observations to support the hypothesis that many forms of biologic responses, including toxic responses, can be extrapolated across mammalian species, including *Homo sapiens,* but the scientific basis of such extrapolation is not established with sufficient rigor to allow broad and definitive generalizations to be made (NRC, 1993b).

One of the most important reasons for species differences in response to chemical exposures is that toxicity is very often a function of chemical metabolism. Differences among animal species, or even among strains of the same species, in metabolic handling of a chemical, are not uncommon and can account for toxicity differences (NRC, 1986). Because in most cases information on a chemical's metabolic profile in humans is lacking (and often unobtainable), identifying the animal species and toxic response most likely to predict the human response

accurately is generally not possible. It has become customary to assume, under these circumstances, that in the absence of clear evidence that a particular toxic response is not relevant to human beings, any observation of toxicity in an animal species is potentially predictive of response in at least some humans (EPA, 1987a). This is not unreasonable, given the great variation among humans in genetic composition, prior sensitizing events, and concurrent exposures to other agents.

As in the case of epidemiologic data, IARC expert panels rank evidence of carcinogenicity from animal studies. It is generally recognized by experts that evidence of carcinogenicity is most convincing when a chemical produces excess malignancies in several species and strains of laboratory animals and in both sexes. The observation that a much higher proportion of treated animals than untreated (control) animals develops malignancies adds weight to the evidence of carcinogenicity as a result of the exposure. At the other extreme, the observation that a chemical produces only a relatively small increase in incidence of mostly benign tumors, at a single site of the body, in a single species and sex of test animal does not make a very convincing case for carcinogenicity, although any excess of tumors raises some concern.

EPA combines human and animal evidence, as shown in Table 4-1, to categorize evidence of carcinogenicity; the agency's evaluations of data on individual carcinogens generally match those of IARC. For noncancer health effects, EPA uses categories like those outlined in Table 4-2. Animal data on other forms of toxicity are generally evaluated in the same way as carcinogenicity data, although this classification looks at hazard identification (qualitative) and dose-response relationships (quantitative) together. No risk or hazard ranking schemes similar to those used for carcinogens have been adopted.

Table 4-1. Categorization of evidence of carcinogenicity

Group		Criteria for Classification
A	Human carcinogen	Sufficient evidence from epidemiologic studies
B	Probable human carcinogen (two subgroups)	Limited evidence from epidemiologic studies and sufficient evidence from animal studies (B1); or inadequate evidence from epidemiologic studies (or no data) and sufficient evidence from animal studies (B2)
C	Possible human carcinogen	Limited evidence from animal studies and no human data
D	Not classifiable as to human carcinogenicity	Inadequate human and animal data or no data
E	Evidence of noncarcinogenicity in humans	No evidence of carcinogenicity from adequate human and animal studies

Source: Adapted from EPA, 1987a.

TABLE 4-2. Weight-of-evidence classification methods for noncancer health effects

Sufficient Evidence

The sufficient-evidence category includes data that collectively provide enough information to judge whether a human developmental hazard could exist within the context of dose, duration, timing, and route of exposure. This category includes both human and experimental-animal evidence.

Sufficient Human Evidence: This category includes data from epidemiologic studies (e.g., case-control and cohort studies) that provide convincing evidence for the scientific community to judge that a causal relationship is or is not supported. A case series in conjunction with strong supporting evidence may also be used. Supporting animal data might or might not be available.

Sufficient Experimental Animal Evidence or Limited Human Data: This category includes data from experimental-animal studies or limited human data that provide convincing evidence for the scientific community to judge whether the potential for developmental toxicity exists. The minimal evidence necessary to judge that a potential hazard exists generally would be data demonstrating an adverse developmental effect in a single appropriate, well-conducted study in a single experimental-animal species. The minimal evidence needed to judge that a potential hazard does not exist would include data from appropriate, well-conducted laboratory-animal studies in several species (at least two) that evaluated a variety of the potential manifestations of developmental toxicity and showed no developmental effects at doses that were minimally toxic to adults.

Insufficient Evidence

This category includes situations for which there is less than the minimal sufficient evidence necessary for assessing the potential for developmental toxicity, such as when no data are available on developmental toxicity, when the available data are from studies in animals or humans that have a limited design (e.g., small numbers, inappropriate dose selection or exposure information, or other uncontrolled factors), when the data are from a single species reported to have no adverse developmental effects, or when the data are limited to information on structure/activity relationships, short-term tests, pharmacokinetics, or metabolic precursors.

Source: EPA, 1987a.

The hazard-identification step of a risk assessment generally concludes with a qualitative narrative of the types of toxic responses, if any, that can be caused by the chemical under review, the strength of the supporting evidence, and the scientific merits of the data and their value for predicting human toxicity. In addition to the epidemiologic and animal data, information on metabolism and on the behavior of the chemical in tissues and cells (i.e., on its mechanism of toxic action) might be evaluated, because clues to the reliability of interspecies extrapolation can often be found here.

Identifying the potential of a chemical to cause particular forms of toxicity in humans does not reveal whether the substance poses a risk in specific exposed populations. The latter determination requires three further analytic steps: emission characterization and exposure assessment (discussed in Chapter 3), dose-response assessment (discussed next), and risk characterization (discussed in Chapter 5).

Dose-Response Assessment

In the United States and many other countries, two forms of dose-response assessment involving extrapolation to low doses are used, depending on the nature of the toxic effect under consideration. One form is used for cancer, the other for toxic effects other than cancer.

Toxic Effects Other Than Cancer

For all types of toxic effects other than cancer, the standard procedure used by regulatory agencies for evaluating the dose-response aspects of toxicity involves identifying the highest exposure among all the available experimental studies at which no toxic effect was observed, the "no-observed-effect level" (NOEL) or "no-observed-adverse-effect level" (NOAEL). The difference between the two values is related to the definition of adverse effect. The NOAEL is the highest exposure at which there is no statistically or biologically significant increase in the frequency of an adverse effect when compared with a control group. A similar value used is the lowest-observed-adverse-effect level (LOAEL), which is the lowest exposure at which there is a significant increase in an observable effect. All are used in a similar fashion relative to the regulatory need. The NOAEL is more conservative than the LOAEL (NRC, 1986).

For example, if a chemical caused signs of liver damage in rats at a dosage of 5 mg/kg per day, but no observable effect at 1 mg/kg per day and no other study indicated adverse effects at 1 mg/kg per day or less, then 5 mg/kg per day would be the LOAEL and 1 mg/kg per day would be the NOAEL under the conditions tested in that study. For human risk assessment, the ratio of the NOAEL to the estimated human dose gives an indication of the margin of safety for the potential risk. In general, the smaller the ratio, the greater the likelihood that some people will be adversely affected by the exposure.

The uncertainty-factor approach is used to set exposure limits for a chemical when there is reason to believe that a safe exposure exists; that is, that its toxic effects are likely to be expressed in a person only if that person's exposure is above some minimum, or threshold. At exposures below the threshold, toxic effects are unlikely. The experimental NOAEL is assumed to approximate the threshold. To establish limits for human exposure, the experimental NOAEL is divided by one or more uncertainty factors, which are intended to account for the uncertainty associated with interspecies and intraspecies extrapolation and other factors. Depending on how close the experimental threshold is thought to be to the exposure of a human population, perhaps modified by the particular conditions of exposure, a larger or smaller uncertainty factor might be required to ensure adequate protection. For example, if the NOAEL is derived from high-quality data in (necessarily limited groups of) humans, even a small safety factor (10 or less) might ensure safety, provided that the NOAEL was derived under conditions of exposure similar to those in the exposed population of interest and the study is otherwise sound. If, however, the NOAEL was derived from a less similar or less reliable laboratory-animal study, a larger uncertainty factor would be required (NRC, 1986).

There is no strong scientific basis for using the same constant uncertainty factor for all situations, but there are strong precedents for the use of some values (NRC, 1986). The regulatory agencies usually require values of 10, 100, or 1,000 in different situations. For example, a factor of 100 is usually applied when the NOAEL is derived from chronic toxicity studies (typically 2-year studies) that are considered to be of high quality and when the purpose is to protect members of the general population who could be exposed daily for a full lifetime (10 to account for interspecies differences and 10 to account for intraspecies differences).

Using the NOAEL/LOAEL/uncertainty-factor procedure yields an estimate of an exposure that is thought to "have a reasonable certainty of no harm." Depending on the regulatory agency involved, the resulting estimate of "safe" exposure can be termed an acceptable daily intake, or ADI (Food and Drug Administration, FDA); a reference dose, or RfD (EPA); or a permissible exposure level, or PEL (Occupational Safety and Health Administration, OSHA). For risk assessments, the dose received by humans is compared with the ADI, RfD, or PEL to determine whether a health risk is likely.

The requirement for uncertainty factors stems in part from the belief that humans could be more sensitive to the toxic effects of a chemical than laboratory animals and the belief that variations in sensitivity are likely to exist within the human population (NRC, 1980a). Those beliefs are plausible, but the magnitudes of interspecies and intraspecies differences for every chemical and toxic end point are not often known. Uncertainty factors are intended to accommodate scientific uncertainty, as well as uncertainties about dose delivered, human variations in sensitivity, and other matters (Dourson and Stara, 1983).

EPA's approaches to risk assessment for chemically induced reproductive and developmental end points rely on the threshold assumption. The EPA (1987a) guidelines for health-risk assessment for suspected developmental toxicants states that, "owing primarily to a lack of understanding of the biological mechanisms underlying developmental toxicity, intra/interspecies differences in the types of developmental events, the influence of maternal effects on the dose-response curve, and whether or not a threshold exists below which no effect will be

produced by an agent," many developmental toxicologists assume a threshold for most developmental effects, because "the embryo is known to have some capacity for repair of the damage or insult" and "most developmental deviations are probably multifactorial."

EPA (1988a,b) later proposed guidelines for assessing male and female reproductive risks that incorporate the threshold default assumption "usually assumed for noncarcinogenic/nonmutagenic health effects," as well as the agency's new RfD approach to deriving acceptable intakes. The RfD is obtained as described above. The total adjustment or uncertainty factor referred to in the proposed guidelines for use in obtaining an RfD from toxicity data "usually ranges" from 10 to 1,000. The adjustment incorporates (as needed) uncertainty factors ("often" 10) for "(1) situations in which the LOAEL must be used because a NOAEL was not established, (2) interspecies extrapolation, and (3) intraspecies adjustment for variable sensitivity among individuals." An additional modifying factor may be used to account for extrapolating between exposure durations (e.g., from acute to subchronic) or for NOAEL-LOAEL inadequacy due to scientific uncertainties in the available database.

EPA's 1992 revision of its guidelines for developmental-toxicity risk assessment state that "human data are preferred for risk assessment" and that the "most relevant information" is provided by good epidemiologic studies. When these data are not available, however, reproductive risk assessment and developmental-agent risk assessment, according to EPA, are based on four key assumptions:

- An agent that causes adverse developmental effects in animals will do so in humans, with sufficient exposure during development, although the types of effects might not be the same in humans as in animals.
- Any significant increase in any of the expressions of developmental toxicants (e.g., death, structural abnormalities, growth alterations, and functional deficits) indicates a likelihood that the agent is a developmental hazard.
- Although the types of effects in humans and animals might not be the same, the use of the most sensitive animal species to estimate human hazards is justified.
- A threshold is assumed in dose-response relationships on the basis of current knowledge, although some experts believe that current science does not fully support this position.

The new guidelines state that "the existence of a NOAEL in an animal study does not prove or disprove the existence or level of a biological threshold." The guidelines also address statistical deficiencies and improvements in the NOAEL-based uncertainty-factor approach (Crump, 1984; Kimmel and Gaylor, 1988; Brown and Erdreich, 1989; Chen and Kodell, 1989; Gaylor, 1989; Kodell et al., 1991a). The guidelines also discuss EPA's plans to move toward a more quantitative "benchmark dose" (BD) for risk assessment for developmental end points "when sufficient data are available"; the BD approach would be consistent with the uncertainty-factor approach now in use (EPA, 1991a). Like the NOAEL and LOAEL, the BD is based on the most sensitive developmental effect observed in the most appropriate or most sensitive mammalian species. It would be derived by modeling the data in the observed range, selecting an incidence rate at a preset low observed response (e.g., 1% or 10%), and determin-

ing the corresponding lower confidence limit on dose that would yield that level of excess response. A BD thus calculated would then be divided by uncertainty factors to derive corresponding acceptable intake (e.g., RfD) values (EPA, 1991a). Thus, the traditional uncertainty-factor approach is retained in the 1991 developmental-toxicity guidelines, as well as in the proposed BD approach. However, the new guidelines are unique, in that they emphasize both the possible effect of interindividual variability in the interpretation of acceptable exposures and the improvements that biologically based models could bring to developmental risk assessment (EPA, 1991a):

> It has generally been assumed that there is a biological threshold for developmental toxicity; however, a threshold for a population of individuals may or may not exist because of other endogenous or exogenous factors that may increase the sensitivity of some individuals in the population. Thus, the addition of a toxicant may result in an increased risk for the population, but not necessarily for all individuals in the population. . . . Models that are biologically based should provide a more accurate estimation of low-dose risk to humans. . . . The Agency is currently supporting several major efforts to develop biologically based dose-response models for developmental toxicity risk assessment that include the consideration of threshold.

CANCER

For some toxic effects, notably cancer, there are reasons to believe either that no threshold for dose-response relationships exists or that, if one does exist, it is very low and cannot be reliably identified (OSTP, 1985; NRC, 1986). This approach is taken on the basis not of human experience with chemical-induced cancer, but rather of radiation-induced cancer in humans and radiologic theory of tissue damage. Risk estimation for carcinogens therefore follows a different procedure from that for noncarcinogens: the relationship between cancer incidence and the dose of a chemical observed in an epidemiologic or experimental study is extrapolated to the lower doses at which humans (e.g., neighboring population) might be exposed (e.g., due to emissions from a plant) to predict an excess lifetime risk of cancer—that is, the added risk of cancer resulting from lifetime exposure to that chemical at a particular dose. In this procedure, there is no "safe" dose with a risk of zero (except at zero dose), although at sufficiently low doses the risk becomes very low and is generally regarded as without public-health significance.

The procedure used by EPA is typical of those used by the other regulatory agencies. The observed relationship between lifetime daily dose and observed tumor incidence is fitted to a mathematical model to predict the incidence at low doses. Several such models are in wide use. The so-called linearized multistage model (LMS) is favored by EPA for this purpose (EPA, 1987a). FDA uses a somewhat different procedure that nevertheless yields a similar result. An important feature of the LMS is that the dose-response curve is linear at low doses, even if it displays nonlinear behavior in the region of observation.

EPA applies a statistical confidence-limit procedure to the linear multistage no-threshold model to generate what is sometimes considered an upper bound on cancer risk. Although the actual risk cannot be known, it is thought that it will not exceed the upper bound, might be

lower, and could be zero. The result of a dose-response assessment for a carcinogen is a potency factor. EPA also uses the term *unit risk factor* for cancer potency. This value is the plausible upper bound on excess lifetime risk of cancer per unit of dose. In the absence of strong evidence to the contrary, it is generally assumed that such a potency factor estimated from animal data can be applied to humans to estimate an upper bound on the human cancer risk associated with lifetime exposure to a specified dosage.

The dose-response step involves considerable uncertainty, because the shape of the dose-response curve at low doses is not derived from empirical observation, but must be inferred from theories that predict the shape of the curve at the low doses anticipated for human exposure. The adoption of linear models is based largely on the science-policy choice that calls for caution in the face of scientific uncertainty. Models that yield lower risks, indeed models incorporating a threshold dose, are plausible for many carcinogens, especially chemicals that do not directly interact with DNA and produce genetic alterations. For example, some chemicals, such as chloroform, are thought to produce cancers in laboratory animals as a result of their cell-killing effects and related stimulation of cell division. However, in the absence of compelling mechanistic data to support such models, regulators are reluctant to use them, because of a fear that risk will be understated. For other substances (e.g., vinyl chloride), evidence shows that the human cancer risk at low doses could be substantially higher than would be estimated by the usual procedures from animal data. Models that yield higher potency estimates at lower doses than the LMS model might also be plausible, but are rarely used (Bailar et al., 1988).

New Trends in Toxicity Assessment

With respect to carcinogenic agents, two types of information are beginning to influence the conduct of risk assessment.

For any given chemical, a multitude of steps can occur between intake and the occurrence of adverse effects. Those events can occur dynamically over an extended period, in some cases decades. One approach to understanding the complex interrelationships is to divide the overall scheme into two pieces, the linkages between exposure and dose and between dose and response. *Pharmacokinetics* has often been used to describe the linkage between exposure (or intake) and dose, and *pharmacodynamics* to describe the linkage between dose and response. Use of the root *pharmaco* (for drug) reflects the origin of those terms. When applied to the study and evaluation of toxic materials, the corresponding terms might more appropriately be *toxicokinetics* and *toxicodynamics*.

Exploration of the use of pharmacokinetic data is especially vigorous. Risk assessors are seeking to understand the quantitative relationships between chemical exposures and target-site doses over a wide range of doses. Because the target-site dose is the ultimate determinant of risk, any nonlinearity in the relationship between administered dose and target-site dose or any quantitative differences in the ratio of the two quantities between humans and test animals could greatly influence the outcome of a risk assessment (which now generally relies on an

assumed proportional relationship between administered and target doses). The problem of obtaining adequate pharmacokinetic data in humans is being attacked by the construction of physiologically based pharmacokinetic (PB-PK) models, whose forms depend on the physiology of humans and test animals, solubilities of chemicals in various tissues, and relative rates of metabolism (NRC, 1989). Several relatively successful attempts at predicting tissue dose in humans and other species have been made with PB-PK modeling, and greater uses of this tool are being encouraged by the regulatory community (NRC, 1987).

A second major trend in risk assessment stems from investigations indicating that some chemicals that increase tumor incidence might do so only indirectly, either by causing first cell-killing and then compensatory cell proliferation or by increasing rates of cell proliferation through mitogenesis. In either case, increasing cell proliferation rates puts cells at increased risk of carcinogenesis from spontaneous mutation. Until a dose of such a carcinogen sufficient to cause the necessary toxicity or intracellular response is reached, no significant risk of cancer can exist. Such carcinogens, or their metabolites, show little or no propensity to damage genes (they are nongenotoxic).

5
RISK CHARACTERIZATION

INTRODUCTION

Characterization of risk is the final step in health risk assessment. This chapter discusses the methods used by the Environmental Protection Agency (EPA) to characterize the public-health risk associated with an emission source. In risk characterization, the assessor takes the exposure information from the exposure-assessment stage (discussed in Chapter 3) and combines it with information from the dose-response assessment stage (discussed in Chapter 4) to determine the likelihood that an emission could cause harm to nearby individuals and populations. The results of this risk characterization are then communicated to the risk manager with an overall assessment of the quality of the information in that analysis. The goal of risk characterization is to provide an understanding of the type and magnitude of an adverse effect that a particular chemical or emission could cause under particular circumstances. The risk manager then makes decisions on the basis of the public-health impact as determined by the risk characterization and other criteria outlined in the appropriate statute.

The elements of risk characterization are discussed here on the basis of several EPA documents, including EPA's *Risk Assessment Guidelines of 1986* (EPA, 1987a); *Guidelines for Exposure Assessment* (EPA, 1992a); a memorandum from Henry Habicht II, deputy administrator of EPA, dated February 26, 1992 (EPA, 1992c) (see Appendix B) (known hereafter as the "risk-characterization memorandum"); and *Risk Assessment Guidance for Superfund* (EPA, 1989a) (the "Superfund document").

ELEMENTS OF RISK CHARACTERIZATION

EPA's risk-characterization step has four elements: generation of a quantitative estimate of risk, qualitative description of uncertainty, presentation of the risk estimate, and communication of the results of risk analysis.

QUANTITATIVE ESTIMATES OF RISK

To determine the likelihood of an adverse effect in an exposed population, quantitative information on exposure—i.e., the dose (determined from the analysis in Chapter 3)—is combined with information on the dose-response relationship (determined from the analysis in Chapter 4). This process is different for carcinogens and for noncarcinogens. For noncarcinogens, the dose estimate is divided by the RfD to obtain a hazard index. If the hazard index is less than 1, the chemical exposure under consideration is regarded as unlikely to lead to adverse health effects. If the hazard index is greater than 1, adverse health effects are more likely and some remedial action is called for. The hazard index is thus not an actual measure of risk; it is a benchmark that can be used to estimate the likelihood of risk.

For carcinogens, excess lifetime risk is calculated by multiplying the dose estimate by a potency factor. The result is a value that represents an upper bound on the probability that lifetime exposure to an agent, under the specified conditions of exposure, will lead to excess cancer risk. This value is usually expressed as a population risk, such as 1×10^{-6}, which means that no more than one in 1 million exposed persons is expected to develop cancer. Risk estimates obtained in this way are *not* scientific estimates of actual cancer risk; they are upper bounds on actual cancer risk that are useful to regulators for setting priorities and for setting exposure limits.

When exposure to more than one agent occurs simultaneously, the cancer risk estimates obtained for each agent can be combined in an additive manner for each route of exposure. Hazard indexes for noncarcinogens may be combined when the agents of concern elicit similar end points of toxicity.

Sometimes, this risk-characterization technique is used to estimate an upper bound on excess lifetime cancer risk to exposed individuals, instead of populations. EPA's *Guidelines for Exposure Assessment* (EPA, 1992a) (not yet implemented) lists some of the questions that should be answered when considering individual versus population risk. These questions are stated by EPA as follows:

Individual Risk

- Are individuals at risk from exposure to the substances under study? Although for substances, such as carcinogens, that are assumed to have no threshold, only a zero dose would result in nonexcess risk for noncarcinogens, this question can often be addressed. In the case of the use of hazard indices, where exposure or doses are compared to a reference dose or some other acceptable level, the risk descriptor would be a statement based on the ratio between the dose incurred and the reference dose.
- To what risk levels are the persons at the highest risk subjected? Who are these people, what are they doing, where do they live, etc., and what might be putting them at this higher risk?
- Can people with a high degree of susceptibility be identified?
- What is the average individual risk?

Population Risk

- How many cases of a particular health effect might be probabilistically estimated for a population of interest during a specified time period?
- For noncarcinogens, what portion of the population exceed the reference dose (RfD), the reference concentration (RfC), or other health concern level? For carcinogens, how many persons are above a certain risk level such as 10^{-6} or a series of risk levels such as 10^{-5}, 10^{-4}, etc.
- How do various subgroups fall within the distributions of exposure, dose, and risk?
- What is the risk for a particular population segment?
- Do any particular subgroups experience a high exposure, dose, or risk?

Description of Uncertainty

Analysis of the uncertainty associated with a health risk estimate involves each step of the risk-assessment process: it brings together the uncertainty in emissions and exposure estimates with that of the toxicity dose-response assessment. Table 5-1 lists the uncertainty issues to be addressed at each step of a health risk assessment. Uncertainty analysis can take place at the time of each of those analyses, but because it affects the eventual risk estimate, it is considered part of the final step of risk assessment—risk characterization.

Several recent documents illustrate EPA's current approach to the analysis of uncertainty associated with health risk assessment, including the Superfund document (EPA, 1989a), the background information document for NESHAPS for radionuclides (EPA, 1989b), the Guidelines for Exposure Assessment (EPA, 1992a), and the risk-characterization memorandum (Appendix B).

Superfund Risk-Assessment Guidance

The Superfund document provides guidance to EPA and other government employees and contractors who are risk assessors, risk-assessment reviewers, remedial project managers, or risk managers involved in Superfund-site cleanup. Section 8.4 of the document "discusses practical approaches to assessing uncertainty in Superfund site risk assessments and describes ways to present key information bearing on the level of confidence in quantitative risk estimates for a site." The document considers three categories of uncertainty associated with site risk assessments: selection of substances, toxicity values, and exposure assessments. Table 5-2 is EPA's uncertainty checklist for Superfund-site risk assessments. Risk assessors are to use the checklist to ensure that they describe adequately the uncertainty in a risk assessment. The document indicates that, although the uncertainty associated with each variable in a risk assessment would ideally be associated with the final risk estimate, a more practical

TABLE 5-1. Uncertainty issues to be addressed in each risk assessment step

A. *Hazard Identification:* What do we know about the capacity of an environmental agent for causing cancer (or other adverse effects) in laboratory animals and in humans?

1. the nature, reliability, and consistency of the particular studies in humans and in laboratory animals;
2. the available information on the mechanistic basis for activity; and
3. experimental animal responses and their relevance to human outcomes.

B. *Dose-Response Assessment:* What do we know about the biological mechanisms and dose-response relationships underlying any effects observed in the laboratory or epidemiology studies providing data for the assessment?

1. relationship between extrapolation models selected and available information on biological mechanisms;
2. how appropriate data sets were selected from those that show the range of possible potencies both in laboratory animals and humans;
3. basis for selecting interspecies dose scaling factors to account for scaling dose from experimental animals to humans; and,
4. correspondence between the expected route(s) of exposure and the exposure route(s) utilized in the hazard studies, as well as the interrelationships of potential effects from different exposure routes.

C. *Exposure Assessment:* What do we know about the paths, patterns, and magnitudes of human exposure and number of persons likely to be exposed?

1. The basis for the values and input parameters used in each exposure scenario. If based on data, information on the quality, purpose, and representatives of the database is needed. If based on assumptions, the source and general logic used to develop the assumption (e.g., monitoring, modeling, analogy, professional judgment) should be described.
2. The major factor or factors (e.g., concentration, body uptake, duration/frequency of exposure) thought to account for the greatest uncertainty in the exposure estimate, due either to sensitivity or lack of data.
3. The link of the exposure information to the risk descriptors. These risk descriptors should include: (1) individual risk including the central tendency and high end portions of the risk distribution, (2) important subgroups of the population such as highly exposed or highly susceptible groups or individuals (if known), and (3) population risk. This issue includes the conservatism or non-conservatism of the scenarios, as indicated by the choice of descriptors. In addition, information that addresses the impact of possible low probability but possibly high consequence events should be addressed.

   For individual risk, information such as the people at highest risk, the risk levels these individuals are subject to, the activities putting them at higher risk, and the average risk for individuals in the population of interest should be addressed. For population risk, information as to the number of cases of a particular health effect that might be probalistically estimated in this population for a specific time period, the portion of the population that are within a specified range of some benchmark level for non-carcinogens; and, for carcinogens, the number of persons above a certain risk level should be included. For subgroups, information as to how exposure and risk impact the various subgroups and the population risk of a particular subgroup should be provided.

Table 5-1 (cont.)

D. *Risk Characterization:* What do other assessors, decision-makers, and the public need to know about the primary conclusions and assumptions, and about the balance between confidence and uncertainty in the assessment? What are the strengths and limitations of the assessment?

1. Numerical estimates should never be separated from the descriptive information that is integral to the risk assessment. For decisionmakers, a complete characterization (key descriptive elements along with numerical estimates) should be retained in all discussions and papers relating to an assessment used in decision-making. Differences in assumptions and uncertainties, coupled with non-scientific considerations called for in various environmental statues, can clearly lead to different risk management decisions in cases with ostensibly identical quantitative risks; i.e., the "number" alone does not determine the decisions.
2. Consideration of alternative approaches involves examining selected plausible options for addressing a given uncertainty. The strengths and weaknesses of each alternative approach and as appropriate, estimates of central tendency and variability (e.g., mean, percentiles, range, variance). The description of the option chosen should include the rationale for the choice, the effect of option selected on the assessment, a comparison with other plausible options, and the potential impacts of new research.

Source: Risk-characterization memorandum (Appendix B).

approach is to describe qualitatively how the uncertainties might be magnified or the estimates of risk biased because of the risk models used. This document is being updated.

Uncertainty Analysis for Radionuclide Risk

EPA undertook a more comprehensive, integrated, quantitative approach to uncertainty characterization in the background document for its environmental impact statement on the National Emission Standards for Hazardous Air Pollutants (NESHAPS) for radionuclides (EPA, 1989b). This document includes an extensive presentation of estimates of fatal cancer risks associated with exposure to radionuclides. The estimates were "intended to be reasonable best estimates of risk; that is, to not significantly underestimate or overestimate risks and be of sufficient accuracy to support decisionmaking" (EPA, 1989b). One chapter of the document, however, provides a detailed analysis of uncertainties in the calculated risks that was undertaken by EPA's Office of Radiation Programs for four selected exposure sites, such as a uranium-mill tailings pile in Washington and an elemental-phosphorus plant in Idaho. The stated reason for the uncertainty analysis was that "quantitative uncertainty analysis can provide results that indicate the likelihood of realizing different risk levels across the range of uncertainty. This type of information is very useful for incorporating acceptable and reasonable confidence levels into decisions" (EPA, 1989b).

TABLE 5-2. EPA guidance for uncertainty analysis in Superfund risk assessments

LIST PHYSICAL SETTING DEFINITION UNCERTAINTIES

- For chemicals not included in the quantitative risk assessment, describe briefly:
  - reason for exclusion (e.g., quality control), and
  - possible consequences of exclusion on risk assessment (e.g., because of widespread contamination, underestimate of risk).
- For the current land uses describe:
  - sources and quality of information, and
  - qualitative confidence level.
- For the future land uses describe:
  - sources and quality of information, and
  - information related to the likelihood of occurrence.
- For each exposure pathway, describe why pathway was selected or not selected for evaluation.
- For each combination of pathways, describe any qualifications regarding the selection of exposure pathways considered to contribute to exposure of the same individual or group of individuals over the same period of time.

CHARACTERIZE MODEL UNCERTAINTIES

- List/summarize the key model assumptions.
- Indicate the potential impact of each on risk:
  - direction (i.e., may over- or underestimate risk); and
  - magnitude (e.g., order of magnitude).

CHARACTERIZE TOXICITY ASSESSMENT UNCERTAINTIES

For each substance carried through the quantitative risk assessment, list uncertainties related to:

- qualitative hazard findings (i.e., potential for human toxicity);
- derivation of toxicity values, e.g.,
  - human or animal data,
  - duration of study (e.g., chronic study used to set subchronic RfD), and
  - any special considerations;
- the potential for synergistic or antagonistic interactions with other substances affecting the same individuals; and
- calculation of lifetime cancer risks on the basis of less-than-lifetime exposures.

For each substance not included in the quantitative risk assessment because of inadequate toxicity information, list:

- possible health effects; and
- possible consequences of exclusion on final risk estimates.

Table 5-2 (cont.)

RISK CHARACTERIZATION

- confidence that the key site-related contaminants were identified and discussion of contaminant concentrations relative to background concentration ranges;
- a description of the various types of cancer and other health risks present at the site (e.g., liver toxicity, neurotoxicity), distinguishing between known effects in humans and those that are predicted to occur based on animal experiments;
- level of confidence in the quantitative toxicity information used to estimate risks and presentation of qualitative information on the toxicity of substances not included in the quantitative assessment;
- level of confidence in the exposure estimates for key exposure pathways and related exposure parameter assumptions;
- the magnitude of the cancer risks and noncancer hazard indices relative to the Superfund site remediation goals in the NCP (e.g., the cancer risk range of 10^{-4} to 10^{-7} and noncancer hazard index of 1.0);
- the major factors driving the site risks (e.g., substances, pathways, and pathway combinations);
- the major factors reducing the certainty in the results and the significance of these uncertainties (e.g., adding risks over several substances and pathways);
- exposed population characteristics; and
- comparison with site-specific health studies, when available.

Source: Adapted from EPA, 1989a.

The EPA uncertainty analysis for radionuclide risks focused on "parameter uncertainty," because it was felt that other sources of uncertainty involving alternative or additional exposure pathways and risk-model structures were "not readily amenable to explicit analysis" (EPA, 1989b). Parameter uncertainties were first modeled as particular probability distributions for each parameter involved in four key components of the radionuclide risk assessments: source terms, atmospheric-dispersion factors, environmental-transport and radionuclide-uptake factors, and risk-conversion (that is, radionuclide-potency) factors. All the distributions pertaining to exposure-related factors were intended to model uncertainty in factor values characteristic of a maximally exposed person. All the distributions pertaining to uptake-related factors were intended to model uncertainty in factor values characteristic of an average individual, except in a set of separate corresponding analyses in which census-based inter-individual variability in home-residence time was incorporated into the analysis, where it was computationally treated as an uncertain parameter.

Monte Carlo methods were used to propagate uncertainty within contamination-uptake-risk models for calculating radionuclide-specific, increased lifetime risks of fatal cancer to an otherwise typical person who is maximally exposed over a lifetime (70 years) or over some shorter period sampled randomly from the distribution used to characterize home-residence time. The resulting characterization obtained for uncertainty in estimated total increased fatal-cancer risk associated with potential maximal exposure to all radionuclides for an exposure scenario involving a uranium-mill tailings pile is shown in Figure 5-1. The horizontal axis in that figure represents increased risk multiplied by 3.5×10^{-6}, which is the geometric mean of the distribution (shown as the solid curve) of risk to an individual maximally exposed for 70

years. (Normalization to the geometric mean value was done simply because all the risk distributions obtained were very close to lognormal.)

The vertical axis in Figure 5-1 represents cumulative probability expressed as a percentage, that is, the probability that the true (but certain) risk is less than or equal to a given, corresponding particular risk value shown on the horizontal axis. The solid horizontal line in the figure corresponds to cumulative probability equal to 50%. The dashed curve in the figure represents estimated risk accounting for less-than-lifetime home residence. In commenting on the substantial difference between the solid and dashed curves for the four types of exposure scenarios considered in this uncertainty analysis, EPA concluded that "it is clear . . . that many moves are to nearby locations," that "we do not believe that including a factor for exposure duration improves the assessment of maximum individual risk," and that "improper application of such a factor can easily lead to erroneous conclusions regarding uncertainties in the risk assessment" (EPA, 1989b).

Presentation of Risk Estimates

Several methods can be used to display health risk estimates. Some of the terms used most often are listed in Table 5-2. The definitions are from the new 1992 exposure guidelines (EPA, 1992a). Any combination of them can be used to display the risk estimate to either the risk manager or the public. The choice of descriptors is often based on legal mandates. In general, the display includes a table indicating the risk estimated for the exposed population by route of exposure.

1992 Exposure-Assessment Guidelines

EPA's 1992 *Guidelines for Exposure Assessment* shows a clear to presentation of hazard-identification, dose-response, and exposure-assessment information that might be useful in future risk assessments. Risk assessors are to examine the judgments made during the process, the constraints of available data, and the state of knowledge. According to EPA, the risk characterization should include (EPA, 1992a)

- the qualitative, weight-of-evidence conclusions about the likelihood that the chemical may pose a specific hazard (or hazards) to human health, the nature and severity of the observed effects, and by what route(s) these effects are seen to occur. These judgments affect both the dose-response and exposure assessments.
- for noncancer effects, a discussion of the dose-response behavior of the critical effect(s), data such as the shapes and slopes of the dose-response curves for the various other toxic end points, and how this information was used to determine the appropriate dose-response assessment techniques; and

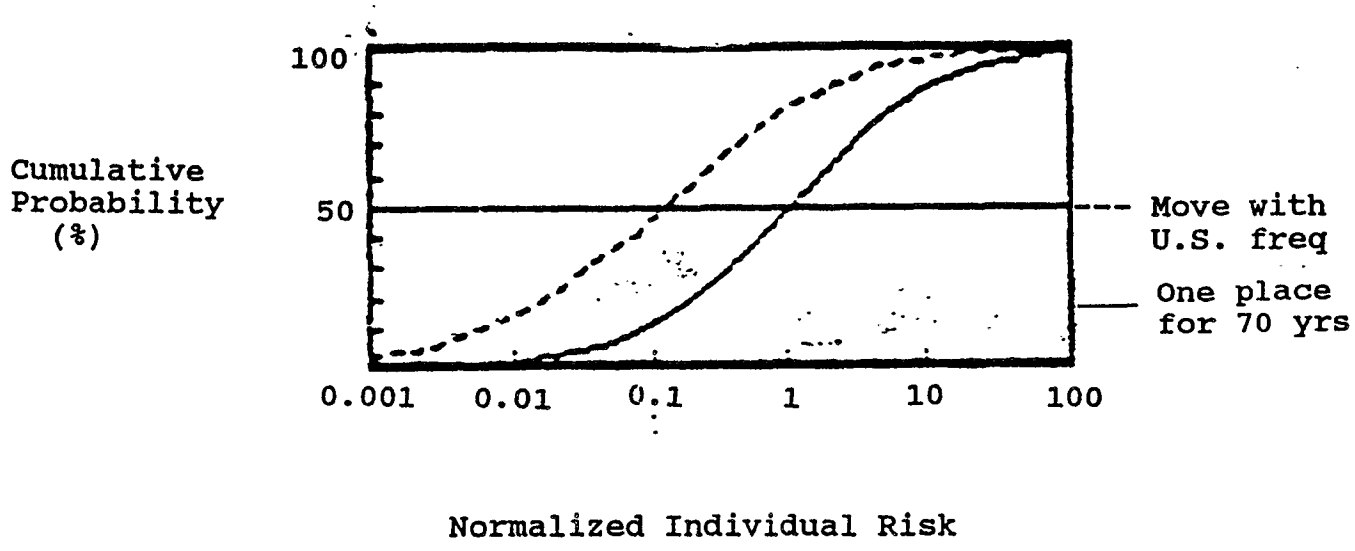


Figure 5-1. Uncertainty in estimated total increased fatal-cancer risk associated with potential maximal exposure to all radionuclides for an exposure scenario involving a uranium-mill tailings pile. Source: Adapted from EPA, 1989b.

- the estimates of the magnitude of the exposure, the route, duration and pattern of the exposure, relevant pharmacokinetics, and the number and characteristics of the population exposed. This information must be compatible with both the hazard identification and dose-response assessments.

The risk-characterization summary should highlight the key points of each step of the risk-assessment process.

RISK-CHARACTERIZATION MEMORANDUM

EPA is in transition on risk characterization. Besides the exposure guidelines described above, the risk-characterization memorandum (Appendix B) provides guidance on risk characterization and uncertainty analysis for EPA risk managers and risk assessors. The memorandum

> addresses a problem that affects public perception regarding the reliability of EPA's scientific assessments and related regulatory decisions. . .Significant information is often omitted as the results of the assessment are passed along in the decision-making process. . . .Often, when risk information is presented to the ultimate decision-maker and to the public, the results have been boiled down to a point estimate of risk. Such "short hand" approaches to risk assessment do not fully convey the range of information considered and used in developing the assessment. In short, informative risk characterizations clarified the scientific basis for EPA decisions, while numbers alone do not give a true picture of the assessment.

A statement attached to the memorandum from the Risk Assessment Council, made up of EPA senior managers, emphasized the following principles:

- *Full Characterization of Risk:* A full and open discussion of uncertainties in the body of each EPA risk assessment, including prominent display of critical uncertainties in the risk characterization. Numerical risk estimates should always be accompanied by descriptive information carefully selected to ensure an objective and balanced characterization of risk in risk assessment reports and regulatory documents.

- *Comparability and Consistency:* Confusion as to the comparability of similar looking (but quite different) risks, for example, the risk estimate for an average individual risk relative to the risk estimate for the most exposed individual, have led to misunderstandings about the relative significance of risks and the protectiveness of risk, reduction action. Therefore, several different descriptors of risk as outlined in the newly revised *Exposure Assessment Guidelines,* should be presented to provide a more complete picture of the risk than available from a single descriptor of risk.

- *Professional Judgment:* There are limits to the degree to which a full characterization of risk may be provided. The degree to which confidence and uncertainty are addressed

depends largely on the scope of the assessment and available sources. So decision-makers and the public are not overwhelmed, only the most significant data and uncertainties need be presented. Further, when special circumstances (e. g., lack of data, extremely complex situations, resource limitations, statutory deadlines) preclude an assessment, such circumstances should be explained.

In implementing that guidance, EPA staff should:

1. Clearly present risk assessment information separate from any non-scientific risk management considerations.

2. Key scientific information on data and methods (e.g., use of animal or human data for extrapolating from high to low doses, use of pharmacokinetics data) must be highlighted, and a statement of confidence in the assessments that identifies all major uncertainties along with comment on their influence on the assessment must be provided.

3. The range of exposures derived from exposure scenarios and on the use of multiple risk descriptors (i.e., central tendency, high end of individual risk, population risk, important subgroups (if known) should be presented.

The risk-characterization memorandum goes through each step of risk assessment and outlines the questions to be answered. These are shown in Table 5-1, which suggests several issues that should be addressed to describe the information in each step fully.

COMMUNICATION OF RISK

Risk communication consists of two parts: communication between the risk assessor and the risk manager and communication between the risk-assessment management team and the public. The risk manager often receives the individual and population risk estimates (generally point estimates but occasionally ranges of these estimates) with only a qualitative description of the uncertainties in each. The general public often receives much less information—only the point estimate or range (without a description of the uncertainty) and the risk manager's decision—although far more is available from published sources or on request. In most regulatory situations, the manager's decision and supporting information are published in the *Federal Register*. In addition, extensive background documents that discuss the risk analysis in much more depth are often available to the public. The public is generally given an opportunity to comment within 30-60 days on the analysis and resulting decision. EPA may adjust a risk assessment on the basis of public comments.

Part II

STRATEGIES FOR IMPROVING RISK ASSESSMENT

Previous chapters have examined the various steps of the health risk-assessment process in the sequence developed by the 1983 Red Book committee. In considering the various steps to risk assessment, the committee observed that several common themes cut across the various stages of risk assessment and arise in criticisms of each individual step. These themes are as follows:

- *Default options.* Is there a set of clear and consistent principles for choosing and departing from default options?
- *Validation.* Has the Environmental Protection Agency (EPA) made a sufficient case that its methods and models for carrying out risk assessments are consistent with current scientific information available?
- *Data needs.* Is enough information available to EPA to generate risk assessments that are protective of public health and are scientifically plausible? What types of information should EPA obtain and how should the information best be used?
- *Uncertainty.* Has EPA taken sufficient account of the need to consider, describe, and make decisions in light of the inevitable uncertainty in risk assessment?
- *Variability.* Has EPA sufficiently considered the extensive variation among individuals in their exposures to toxic substances and in their susceptibilities to cancer and other health effects?
- *Aggregation.* Is EPA appropriately addressing the possibility of interactions among pollutants in their effects on human health, and addressing the consideration of multiple exposure pathways and multiple adverse health effects?

The "Red Book" paradigm should be supplemented by applying a cross-cutting approach that uses those themes. Such an approach could ameliorate the following problems in risk assessment as it is currently practiced within the agency:

- The differing opinions in the scientific community on the merits of particular scientific evidence and the resulting lack of credibility caused by periodic revisions of particular "risk numbers" (e.g., those for dioxin).
- The reluctance to incorporate new scientific information into risk assessments when it might (erroneously) appear to increase uncertainty.
- The incompatibility of various inputs to risk characterization, e.g., dose estimates in units

that cannot be combined with more sophisticated dose-response evaluations, or hazard-identification evidence that cannot readily be integrated into potency assessment.

- The emphasis on theoretical modeling over measurement.
- The production of risk assessments that are either insufficiently informative or too detailed for the needs of risk managers, and the related problem of lack of clear signals to guide risk-assessment research.

Considering the six cross-cutting themes in the planning and analysis of risk assessment will not solve the problems of risk assessment by itself. Indeed, too much emphasis on a cross-cutting vision of risk assessment might create unanticipated problems. On balance, however, the view of risk assessment proposed in Chapters 6-11 will serve two important purposes: it will give the individual cross-cutting themes a more prominent place in the risk-assessment process, and it will encourage the gradual evolution of attempts to improve risk assessment from its current, somewhat piecemeal orientation to a more holistic one, with the goal of improving the precision, comprehensibility, and usefulness for regulatory decision-making of the *entire* risk-assessment process. Whatever conceptual framework is used, the committee believes that EPA must develop principles for choosing default options and for judging when and how to depart from them. This controversial issue is described in the next section.

THE NEED FOR RISK-ASSESSMENT PRINCIPLES

Our scientific knowledge of hazardous air pollutants has numerous gaps. Hence, there are many uncertainties in the health risk assessments of those pollutants. Some of these can be referred to as model uncertainties—for example, uncertainties regarding dose-response model choices due to a lack of knowledge about the mechanisms by which hazardous air pollutants elicit toxicity. As discussed more fully in Chapter 6, EPA has developed "default options" to use when such uncertainties arise. These options are used in the absence of convincing scientific information on which of several competing models and theories is correct. The options are not rules that bind the agency; rather, they constitute guidelines from which the agency may depart when evaluating the risks posed by a specific substance. The agency may also change the guidelines as scientific knowledge accumulates.

The committee, as discussed in Chapter 6, believes that EPA has acted reasonably in electing to issue default options. Without uniform guidelines, there is a danger that the models used in risk assessment will be selected on an ad hoc basis, according to whether regulating a substance is thought to be politically feasible or according to other parochial concerns. In addition, guidelines can provide a predictable and consistent structure for risk assessment.

The committee believes that only the description of default options in a risk assessment is not adequate. We believe that EPA should have principles for choosing default options and for judging when and how to depart from them. Without such principles, departures from defaults could be ad hoc, thereby undercutting the purpose of the default options. Neither the agency nor interested parties would have any guidance about the quality or quantity of evidence

necessary to persuade the agency to depart from the default options or the point(s) in the process at which to present that evidence.

Moreover, without an underlying set of principles, EPA and the public will have no way to judge the wisdom of the default options themselves. The individual default options inevitably vary in their scientific basis, foundation in empirical data, degree of conservatism, plausibility, simplicity, transparency, and other attributes. If defaults were chosen without conscious reference to these or other attributes, EPA would be unable to judge the extent to which they fulfill the desired attributes. Nor could the agency make intelligent and consistent judgment about when and how to add new default options when "missing defaults" are identified. In addition, the policies that underlie EPA's choice of risk-assessment methods would not be clear to the public and Congress—for example, it would be unclear whether EPA places the highest value on protecting public health, on generating scientifically accurate estimates, or on other concerns.

The committee has identified a number of objectives that should be taken into account when considering principles for choosing and departing from default options: protecting the public health, ensuring scientific validity, minimizing serious errors in estimating risks, maximizing incentives for research, creating an orderly and predictable process, and fostering openness and trustworthiness. There might be additional relevant criteria as well.

The choice of principles inevitably involves choosing how to balance such objectives. For instance, the most open process might not be the one that yields the result most likely to be scientifically valid. Similarly, the goal of minimizing errors in estimation might conflict with that of protecting the public health, inasmuch as (given the pervasiveness of uncertainty) achievement of the latter objective might involve accepting the possibility that a given risk assessment will overestimate the risk.

The committee therefore found it difficult to agree on what principles EPA should adopt. For example, the committee debated whether EPA should base its practices on "plausible conservatism"—that is, on attempting to use models that have support in the scientific community and that tend to minimize the possibility that risk estimates generated by these models will significantly underestimate true risks. The committee also discussed whether EPA instead should attempt as much as possible to base its practices on calculating the risk estimate most likely to be true in the light of current scientific knowledge. After extensive discussion, no consensus was reached on this issue.

The committee also concluded that the choice of principles to guide risk assessment, although it requires a knowledge of science and scientific judgment, ultimately depends on policy judgments, and thus is not an issue for specific consideration by the committee, even if it could agree on the substance of specific recommendations. The choice reflects decisions about how scientific data and inferences should be used in the risk-assessment process, not about which data are correct or about what inferences should be drawn from those data. Thus, the selection of principles inevitably involves choices among competing values and among competing judgments about how best to respond to uncertainty.

Many members contended that the committee ought not attempt to recommend principles, but should leave their formulation to the policy process. They concluded that weighing

societal values is properly left to those who have been chosen, directly or indirectly, to represent the public. Indeed, in the view of these members, any recommendation by the committee would give the false impression that the choice of principles is ultimately an issue of science; noting the sharp differentiation that Congress made between the tasks of this committee and those of the Risk Assessment and Management Commission established by Section 303 of the Clean Air Act Amendments of 1990. That commission, rather than this committee, appears to have been intended to address issues of policy.

Other members contended that the committee should attempt to recommend principles. They urged that the choice of risk-assessment principles is one of the most important decisions to be made in risk assessment and one on which risk assessment experts, because of their expertise on the scientific issues related to the choice, ought to make themselves heard. They believe that the choice of principles is no more policy-laden than many other issues addressed by the committee, and that the decision not to recommend principles is itself a policy choice. They also note that the scientific elements involved in making the choice distinguish the selection of principles from other pure "policy" issues that the committee agreed not to address such as the use of cost-benefit methods or the implications of the psychosocial dimensions of risk perception.

The committee has decided not to recommend principles in its report. Instead, it has included in Appendix N papers by three of its members that offer various perspectives on the issue. One paper, by Adam Finkel, urges that EPA should strive to advance scientific consensus while minimizing serious errors of risk underestimation, by adopting an approach of "plausible conservatism." The other, by Roger McClellan and Warner North, argues that EPA should promote risk assessments that reflect current scientific understanding. Those perspectives are not intended to reflect the total range of opinion among committee members on the subject, but are presented to illustrate the issues involved.

Reporting Risk Assessments

As already mentioned, uncertainties are pervasive in risk assessment. When uncertainty concerns the magnitude of a physical quantity that can be measured or inferred from assumptions (e.g., ambient concentration), it can often be quantified, as Chapter 9 suggests.

Model uncertainties result from an inability to determine which scientific theory is correct or what assumptions should be used to derive risk estimates. Such uncertainties cannot be quantified on the basis of data. Any expression of probability, whether qualitative (e.g., a scientist's statement that a threshold is likely) or quantitative (e.g., a scientist's statement that there is a 90% probability of a threshold), is likely to be subjective. Subjective quantitative probabilities could be useful in conveying the judgments of individual scientists to risk managers and to the public, but the process of assessing subjective probabilities is difficult and essentially untried in a regulatory context. Substantial disagreement and misunderstanding about the reliability of quantitative probabilities could occur, especially if their basis is not set forth clearly and in detail.

In the face of important model uncertainties, it may still be undesirable to reduce a risk characterization to a single number, or even to a range of numbers intended to portray uncertainty. Instead, EPA should consider giving risk managers risk characterizations that are both qualitative and quantitative and both verbal and mathematical.

If EPA takes this route, quantitative assessments provided to risk managers should be based on the principles selected by EPA. EPA might choose to require that a risk assessment be accompanied by a statement describing alternative assumptions presented to the agency that, although they do not meet the principles selected by EPA for use in the risk characterization, satisfy some lesser test (e.g., plausibility). For example, EPA generally assumes that no threshold exists for carcinogenicity and calculates cancer potency using the linearized multistage model as the default. Commenters to the agency on a specific substance might attempt to show that there is a threshold for that substance on the basis of what is known about its mechanism of action. If the threshold can be demonstrated in a manner that is satisfactory under the agency's risk-assessment principles, the risk characterization would be based on the threshold assumption. If such a demonstration cannot be made, then the risk characterization would be based on the no-threshold assumption; but if the threshold assumption were found to be plausible, the risk manager might be informed of its existence as a plausible assumption, its rationale, and its effect on the risk estimate. In this way, risk assessors would receive both qualitative and quantitative information relevant to characterizing the uncertainty associated with the risk estimate.

THE ITERATIVE APPROACH

One strategy component that deserves emphasis is the need for iteration. Neither the resources nor the necessary scientific data exist to perform a full-scale risk assessment on each of the 189 chemicals listed as hazardous air pollutants by Section 112 of the Clean Air Act. Nor, in many cases, is such an assessment needed. Some of the chemicals are unlikely to pose more than a de minimis (trivial) risk once the maximum available control technology is applied to their sources as required by Section 112. Moreover, most sources of Section 112 pollutants emit more than one such pollutant, and control technology for Section 112 pollutants is rarely pollutant-specific. Therefore, there might not be much incentive for industry to petition EPA to remove substances from Section 112's list (or much need for EPA to devote its resources to carrying out risk assessments in response to such petitions).

An iterative approach to risk assessment would start with relatively inexpensive screening techniques and move to more resource-intensive levels of data-gathering, model construction, and model application as the particular situation warranted. To guard against the possibility of underestimating risk, screening techniques must be constructed that err on the side of caution when there is uncertainty. (As discussed in Chapter 12, the committee has some doubts about whether EPA's current screening techniques are so constructed). The results of such screening should be used to set priorities for gathering further data and applying successively more complex techniques. These techniques should then be used to the extent necessary to make a

judgment. In Chapter 7, the kinds of data that should be obtained at each stage of such an iterative process are described. The result would be a process that yields the risk-management decisions required by the Clean Air Act and that provides incentives for further research without the need for costly case-by-case evaluations of individual chemicals. Use of an iterative approach can improve the scientific basis of risk-assessment decisions and account for risk-management concerns, such as the level of protection and resource constraints.

6
DEFAULT OPTIONS

EPA's risk-assessment practices rest heavily on "inference guidelines" or, as they are often called, "default options." These options are generic approaches, based on general scientific knowledge and policy judgment, that are applied to various elements of the risk-assessment process when the correct scientific model is unknown or uncertain. The 1983 NRC report *Risk Assessment in the Federal Government: Managing the Process* defined *default option* as "the option chosen on the basis of risk assessment policy that appears to be the best choice in the absence of data to the contrary" (NRC, 1983a, p. 63). Default options are not rules that bind the agency; rather, as the alternative term *inference guidelines* implies, the agency may depart from them in evaluating the risks posed by a specific substance when it believes this to be appropriate. In this chapter, we discuss EPA's practice of adopting guidelines containing default options and departing from them in specific cases.

ADOPTION OF GUIDELINES

As our discussion of risk assessment has made clear, current knowledge of carcinogenesis, although rapidly advancing, still contains many important gaps. For instance, for most carcinogens, we do not know the complete relationship between the dose of a carcinogen and the risk it poses. Thus, when there is evidence of a carcinogenic effect at a high concentration (for instance, in the workplace or in animal testing), we do not know for certain how strong the effect (if any) would be at the lower concentrations typically found in the environment. Similarly, we do not know how much importance to attach to experiments that show that exposure to a substance causes only benign tumors in animals or how to adjust for metabolic differences between animals and humans in calculating the carcinogenic potency of a chemical.

Other uncertainties are not peculiar to carcinogenesis, but are characteristic of many aspects of risk assessment. For example, calculating the doses received by individuals might require knowledge of the relationship between emission of a substance by a source and the ambient concentration of that substance at a particular place and time. It is impossible to install a monitor at every place where people might be exposed; moreover, monitoring results are subject to error. Thus, regulators attempt to use air-quality models to predict ambient concentrations. But because our knowledge of atmospheric processes is imperfect and the data

needed to use the models cannot always be obtained, the predictions from atmospheric-transport models can differ substantially from measured ambient concentrations (NRC, 1991a).

In time, we hope, our knowledge and data will improve. Indeed, we believe that EPA and other government agencies must engage in scientific research and be receptive to the results of sound scientific research conducted by others. In the meantime, decisions about regulating hazardous air pollutants must be made under conditions of uncertainty. It is vital that the risk-assessment process handle uncertainties in a predictable way that is scientifically defensible, consistent with the agency's statutory mission, and responsive to the needs of decision-makers.

These uncertainties, as we explain further in Chapter 9, are of two major types. One type, which we call *parameter uncertainty*, is caused by our inability to determine accurately the values of key inputs to scientific models, such as emissions, ambient concentrations, and rates of metabolic action. The second type, *model uncertainty*, is caused by gaps in our knowledge of mechanisms of exposure and toxicity—gaps that make it impossible to know for certain which of several competing models is correct. For instance, as mentioned above, we often do not know whether a threshold may exist below which a dose of a carcinogen will not result in an adverse effect. As we discuss in Chapter 9, model uncertainties, unlike parameter uncertainties, are often difficult to quantify.

The Red Book recommended that model uncertainties be handled through the development of uniform inference guidelines for the use of federal regulatory agencies in the risk-assessment process. Such guidelines would structure the interpretation of scientific and technical information relevant to the assessment of health risks. The guidelines, the report urged, should not be rigid, but instead should allow flexibility to consider unique scientific evidence in particular instances.

The Red Book described the advantages of such guidelines as follows (pp. 7-8):

> The use of uniform guidelines would promote clarity, completeness, and consistency in risk assessment; would clarify the relative roles of scientific and other factors in risk assessment policy; would help to ensure that assessments reflect the latest scientific understanding; and would enable regulated parties to anticipate government decisions. In addition, adherence to inference guidelines will aid in maintaining the distinction between risk assessment and risk management.

This committee believes that those considerations continue to be valid. In particular, we stress the importance of inference guidelines as a way of keeping risk assessment and risk management from unduly influencing each other. Without uniform guidelines, risk assessments might be manipulated on an ad hoc basis according to whether regulating a substance is thought to be politically feasible. In addition, we believe that inference guidelines can provide a predictable and consistent structure for risk assessment and that a statement of guidelines forces an agency to articulate publicly its approach to model uncertainty.

Like the committee that produced the 1983 NRC report, we recognize that there is an inevitable interplay between risk assessment and risk management. As the 1983 report states (pp. 76, 81), "risk assessment must always include policy, as well as science," and "guidelines must include both scientific knowledge and policy judgments." Any choice of defaults, or the decision not to have defaults at all, therefore amounts to a policy decision. Indeed, without a

policy decision, the report stated, risk-assessment guidelines could do no more than "state the scientifically plausible inference options for each risk assessment component without attempting to select or even suggest a preferred inference option" (NRC, 1983a, p. 77). Such guidelines would be virtually useless. The report urged that risk-assessment guidelines include risk-assessment policy and explicitly distinguish between scientific knowledge and risk-assessment policy to keep policy decisions from being disguised as scientific conclusions (NRC, 1983a, p. 7). That report urged that for consistency, policy judgments related to risk assessment ought to be based on a common principle or principles.

We believe that EPA acted reasonably in electing to issue *Guidelines for Carcinogen Risk Assessment* (EPA, 1986a). Those guidelines set out policy judgments about the accommodation of model uncertainties that are used to assess risk in the absence of a clear demonstration that a particular theory or model should be used.

For instance, the default options indicate that, in assessing the magnitude of risk to humans associated with low doses of a substance, "in the absence of adequate information to the contrary, the linearized multistage procedure will be employed" (EPA, 1986a, p. 33997). The linearized multistage procedure implies low-dose linearity. At low doses, if the dose is reduced by, say, a factor of 1,000, the risk is also reduced by a factor of 1,000; dose is linearly related to risk. Departure from this default option is allowed, under EPA's current guidelines, if there is "adequate evidence" that the mechanism through which the substance is carcinogenic is more consistent with a different model—for instance, that there is a threshold below which exposure is not associated with a risk. Thus, the default option in guiding a decisionmaker, in the absence of evidence to the contrary, assigns the burden of persuasion to those who wish to show that the linearized multistage procedure should not be used. Similar default options cover such important issues as the calculation of effective dose, the treatment of benign tumors, and the procedure for scaling animal-test results to estimates of potency in humans.

Some default options are concerned with issues of extrapolation—from laboratory animals to humans, from large to small exposures (or doses), from intermittent to chronic lifetime exposures, and from route to route (as from ingestion to inhalation). That is because few chemicals have been shown in epidemiologic studies to cause measurable numbers of human cancers directly, and epidemiologic data on only a few of these are sufficient to support quantitative estimates of human epidemiologic cancer risk. In the absence of adequate human data, it is necessary to use laboratory animals as surrogates for humans.

One advantage of guidelines, as already noted, is that they can articulate both the agency's choice of individual default options and its rationale for choosing all of the options. EPA's guidelines set out individual options but do not do so with ideal clarity. Nor has the agency explicitly articulated the scientific and policy bases for its options. Hence, there might be disagreement about precisely what the agency's default options are and the rationales for these options. We attempt here to identify the most important of the options (numbered points in the 1986 guidelines are cited):

- Laboratory animals are a surrogate for humans in assessing cancer risks; positive cancer-bioassay results in laboratory animals are taken as evidence of a chemical's cancer-causing potential in humans (IV).
- Humans are as sensitive as the most sensitive animal species, strain, or sex evaluated in a bioassay with appropriate study-design characteristics (III.A.1).
- Agents that are positive in long-term animal experiments and also show evidence of promoting or cocarcinogenic activity should be considered as complete carcinogens (II.B.6).
- Benign tumors are surrogates for malignant tumors, so benign and malignant tumors are added in evaluating whether a chemical is carcinogenic and in assessing its potency (III.A.1 and IV.B.1).
- Chemicals act like radiation at low exposures (doses) in inducing cancer; i.e., intake of even one molecule of a chemical has an associated probability for cancer induction that can be calculated, so the appropriate model for relating exposure-response relationships is the linearized multistage model (III.A.2).
- Important biological parameters, including the rate of metabolism of chemicals, in humans and laboratory animals are related to body surface area. When extrapolating metabolic data from laboratory animals to humans, one may use the relationship of surface area in the test species to that in humans in modifying the laboratory animal data (III.A.3).
- A given unit of intake of a chemical has the same effect, regardless of the time of its intake; chemical intake is integrated over time, irrespective of intake rate and duration (III.B).
- Individual chemicals act independently of other chemicals in inducing cancer when multiple chemicals are taken into the body; when assessing the risks associated with exposures to mixtures of chemicals, one treats the risks additively (III.C.2).

EPA has never articulated the policy basis for those options. As we discuss in the previous introductory section (Part II), the agency should choose and explain the principles underlying its choices to avoid the dangers of ad hoc decision-making. The agency's choices are for the most part intended to be conservative—that is, they represent an implicit choice by the agency, in dealing with competing plausible assumptions, to use (as default options) the assumptions that lead to risk estimates that, although plausible, are believed to be more likely to overestimate than to underestimate the risk to human health and the environment. EPA's risk estimates thus are intended to reflect the upper region of the range of risks suggested by current scientific knowledge.

EPA appears to use conservative assumptions to implement Congress's authorization in several statutes, including the Clean Air Act, for the agency to undertake preventive action in the face of scientific uncertainty (see, e.g., *Ethyl v. EPA*, 541 F.2d 1 (D.C. Cir.) (en banc), *certiorari denied* 426 U.S. 941 (1976), ratified by Section 401 of the Clean Air Act Amendments of 1977) and to set standards that include a precautionary margin of safety against unknown effects and errors in calculating risks (see *Environmental Defense Fund v. EPA*, 598 F.2d 62, 70 (D.C. Cir. 1978) and *Natural Resources Defense Council v. EPA*, 824 F.2d 1146, 1165 (en banc) (D.C. Cir. 1987)).

EPA's choice of defaults has been controversial. We note, though, that some of the arguments about EPA's practices are directed less at conservatism than at the means of implementation that the agency has adopted. We believe that the iterative approach recommended in the previous chapter combined with quantitative uncertainty analysis will improve the agency's practices regardless of the degree of conservatism chosen by the agency. We also note that with an iterative approach, the agency must use relatively conservative models in performing screening estimates designed to indicate whether a pollutant is worthy of further analysis and comprehensive risk assessment. Such estimates are intended to obviate the detailed assessment of risks that can with a high degree of confidence be deemed acceptable or de minimis (trivial). By definition, therefore, screening analyses must be sufficiently conservative to make sure that a pollutant that could pose dangers to health or welfare will receive full scrutiny.

Over time, the choice of defaults should have decreasing impact on regulatory decision-making. As scientific knowledge increases, uncertainty diminishes. Better data and increased understanding of biological mechanisms should enable risk assessments that are less dependent on default assumptions and more accurate as predictions of human risk.

In evaluating EPA's risk-assessment methods, we are aware that the agency's guidelines, to use the terminology of the earlier NRC report, are in part statements of science policy, rather than purely statements of scientific fact. The guideline cited above dealing with extrapolation of high doses to low doses is illustrative. The guideline is not a claim that it is known that the relationship between dose and response is linear; that the true relationship between dose and response is uncertain and could be nonlinear is readily acknowledged. Rather, the guideline is based (1) on the scientific conclusion that the linear model has substantial support in current data and biologic theory and that no alternative model has sufficient support to warrant departure from the linear model for most chemicals identified as carcinogens; (2) on the further scientific conclusion that the linear model is more conservative than most alternative plausible models; and (3) on the policy judgment that a conservative model should be chosen when there is model uncertainty.

DEPARTURES FROM DEFAULT OPTIONS

Agency policies should encourage further scientific research. Risk assessors and managers must be receptive to new scientific information about the character and magnitude of the toxic effects of a chemical substance. Putting this receptivity into practice, though, has proved difficult. The 1983 NRC report criticized how agencies had implemented their guidelines. The report noted that "the application of inference options to specific risk assessments has been marked by a general lack of explicitness" and that that made it "difficult to know whether assessors adhere to guidelines" (NRC, 1983a, p. 79). The NRC report recognized the need to prevent ad hoc and undocumented departures from guidelines in specific risk assessments. But the NRC report made it clear that well-designed guidelines "should permit acceptance of new evidence that differs from what was previously perceived as the general case, when scientifi-

cally justifiable." NRC urged a recognition of the need for a tradeoff between flexibility on the one hand and predictability and consistency on the other (NRC, 1983a, p. 81).

The NRC advocated that agencies seek a middle path between inflexibility and ad hoc judgments, but steering this course is difficult. Consistency and predictability are served if an agency sets out criteria for departing from its guidelines. If such criteria are themselves too rigidly applied, the guidelines could ossify into inflexible rules; but without such criteria, the guidelines could be subverted at will with the potential for political manipulation of risk assessment.

NRC's approach requires that agencies regard their inference options not as binding rules, but rather as guidelines that are to be followed unless a sufficient showing is made. In the decade since the NRC report, EPA has never articulated clearly its criteria for a departure. We believe that a structured approach would give better guidance to the scientific community and to the public and would ensure both that the default options are set aside only when there is a valid scientific reason for doing so and that decisions to set aside defaults are scientifically credible and receive public acceptance.

EPA's practice appears to be to allow departure in a specific case when it ascertains that there is a consensus among knowledgeable scientists that the available scientific evidence justifies departure from the default option. The agency apparently considers both the quality of the data submitted and the robustness of the theory that is used to justify the departure.

EPA needs to be more precise in describing the kind and strength of evidence that it will require to depart from a default option. Because the decision as to the evidentiary burden to be required is ultimately one of policy, and because we could not reach agreement on proposed language to implement such a standard (see Appendixes N-1 and N-2), we do not urge any particular standard; moreover, we are conscious of the difficulties of capturing the nuances of judgment in any verbal formula that will not be open to misinterpretation.

We believe that the agency must continue to rely on its Science Advisory Board (SAB) and other expert bodies to determine when departing from a default option is warranted according to default options EPA will develop. EPA has increasingly used peer review and workshops as a way to ensure that it carefully considers the propriety of departing from a default. These and other devices should continue to ensure broad peer and scientific participation to guarantee, as much as possible, that the agency's risk-assessment decisions are made with access to the best science available.

We note that here, too, EPA has a difficult path to tread. EPA has been criticized for delay in deciding whether to depart from default options. Increased procedural formality raises the possibility of further delays, especially in a period of budgetary stringency such as EPA can expect to face for some time. It is likely that EPA will be cutting back on hiring personnel at the salary ranks necessary to attract scientists with the needed experience and training to judge whether departure from a default option is justifiable. Congress ought to be aware of the need for greater agency resources to carry out the mandates of the Clean Air Act and similar legislation.

Even if a default option is not set aside, we believe that decision-makers ought to be informed in a narrative way of any specific information suggesting that, in specific cases,

alternatives to the default options might have equal or greater scientific support, and believe that the characterization of risk should include a discussion of the effect of the alternative options on risk estimates.

CURRENT EPA PRACTICE IN DEPARTING FROM DEFAULT OPTIONS

As discussed above, EPA needs simultaneously to be receptive to evidence indicating the need to depart from a default option and to be careful that it departs from a default in a specific case only when a departure is justifiable. In addition, the agency needs to follow a process that allows peer participation and review.

We discuss below some of the cases in which EPA has addressed the issue of whether to depart from default options. In each of these cases, EPA decisions to depart from default options lessened its estimate of the risk; however, it is important to note that new scientific data could also increase the estimate of risk above that reached by using the default options.

EXAMPLE 1: USE OF ANIMAL-CANCER BIOASSAY DATA

The example that follows illustrates a departure from the two default options that: (1) positive animal-bioassay results for cancer induction are sufficient proof of cancer hazard in humans; and (2) that humans are at least as sensitive as the most sensitive responding animal species. It involves induction of kidney cancer in male laboratory rats by a number of chemicals—most important, 1,4-dichlorobenzene, hexachloroethane, isophorone, tetrachloroethylene, dimethyl methyl phosphorate, d-limonene, pentachloroethane, and unleaded gasoline (EPA, 1991d). The first four have been classified as hazardous air pollutants by the Clean Air Act Amendments of 1990.

Male rats exposed to those chemicals develop dose-related kidney cancer; the highest incidence is usually 25% or less. The tumors do not occur in other organs or other species or in female rats. Because of the economic importance of several of the compounds and unleaded gasoline, extensive studies were conducted to understand the mechanisms involved in the development of the tumors. The studies suggested that a special mechanism was responsible for the tumors in male rats. When the chemicals in question are inhaled by male rats, the chemicals, or products of their metabolism, reach the bloodstream and form complexes with a specific protein, alpha-2μ-globulin, that is produced in the male rat liver and removed from the blood by the kidneys. As the complex is cleared from the blood by the kidneys, it accumulates there in the form of hyaline droplets, which lead to the development of kidney disease characterized by cell death, cast formation, mineralization, and hyperplasia. This accumulation, as well as statistically significant increases in tumors that result from exposure to the chemicals, occurs only in male rats.

In contrast, female rats, which do not have the same concentrations of alpha-2μ-globulin protein, do not develop statistically significant increases tumors as a result of exposure. Similarly, the protein is not present in detectable quantities in humans, so no risk of kidney-cancer development by this mechanism would be expected in humans exposed to the chemicals in question. It was therefore suggested that, inasmuch as a special mechanism not found in humans seemed to be responsible for the tumors, EPA ought to depart in this case from its default option that a substance that is carcinogenic in animals is also a human carcinogen. In response, EPA (1991d) evaluated the evidence of production of kidney tumors in male rats by chemicals inducing alpha-2μ-globulin accumulation (CIGAs), such as those in question. EPA's review suggested that kidney cancer in male rats from exposure to CIGAs is due only to the kidney disease that CIGAs cause through accumulation of alpha 2μ-globulin. For instance, EPA noted, the CIGAs are not known to react with DNA and are generally negative in short-term tests for genotoxicity. In contrast, classical kidney carcinogens (or their active metabolites) are usually electrophilic species that bind covalently to macromolecules and form DNA adducts. With the classical kidney carcinogens, which presumably are carcinogenic in both laboratory animals and humans, the kidney carcinogenesis is presumed to result from the interaction of the compounds or their metabolites with DNA. Classical kidney carcinogens, such as dimethylnitrosamine, induce renal tubule cancer in laboratory animals at a high incidence in both sexes after short periods of exposure, with a clear increase in kidney tumor incidence with increased dose. Thus, the classical kidney carcinogens and CIGAs appear to act via different mechanisms.

After reviewing the data, EPA (1991d) provided specific decision criteria for categorizing a chemical as a CIGA. A substance may be so classified only if it meets all the decision criteria, and classification of a chemical as a CIGA does not keep it from being considered as a carcinogen because of other modes of action. In that way, the agency precisely tailored its proposed departure from default options. EPA concluded that renal tubule tumors in male rats attributable solely to chemically induced alpha-2μ-globulin accumulation should not be used for human-cancer hazard identification or for dose-response extrapolations. Furthermore, EPA noted that even in the absence of renal tubule tumors in the male rat, if the lesions of alpha-2μ-globulin syndrome are present, the associated nephropathy in male rats should not contribute to determinations of noncarcinogenic hazard or risk.

EPA's documents reviewed and synthesized the available scientific information in a document that was then presented to peers in a public meeting, reviewed by the SAB's Environmental Health Committee and later endorsed by the SAB Executive Committee, and transmitted to the administrator (EPA, 1991d). Transmission to the administrator was accompanied by endorsement by the SAB that the document outlined a scientifically sound policy for departing from the default option for this specific class of compounds. This policy has been generally supported by the scientific community. However, it is noteworthy that some researchers (see, e.g., Melnick, 1993) believe that another mechanism to explain all of the observed data is equally or more plausible than the one EPA endorsed. Alpha-2μ-globulin may be a carrier protein that transports certain chemicals to the kidney, where toxic metabolites can be released; this mechanism defines alpha-2μ-globulin accumulation as an *indicator*,

rather than the *cause* of renal toxicity. If so, humans may have other carrier proteins that could transport toxins to the kidney and cause toxicity or carcinogenicity in the absence of protein droplet information, and the assumption that the rat studies are irrevelant to humans might therefore be erroneous.

EXAMPLE 2: LINKAGES BETWEEN EXPOSURE, DOSE, AND RESPONSE

In the previous example, a departure from default options occurred at the hazard-identification stage. As discussed in examples 2 and 3, such departures can also be used to refine the unit risk estimate of a carcinogen.

Calculating the unit risk through quantitative risk assessment requires an understanding of the relationship between exposure to a substance and response. One part of this relationship involves the link between exposure (that is, intake of a substance) and dose (that is, the amount of the substance, or harmful metabolites, that is taken up by bodily organs). However, that understanding is incomplete. EPA's default options assume that all species are equally sensitive to a given target-tissue dose of the toxicant or its metabolites. The surface-to-area ratios in the test species and humans are used as the key to relating the dose received by the test species to the dose that would cause similar effects in humans (see pp. 6-7, III.A.3). As the following examples show, however, evidence can sometimes support departing from this default option.

METHYLENE CHLORIDE

Epidemiological studies on whether exposure to methylene chloride causes cancer in humans have produced equivocal results. Thus, assessment of methylene chloride's carcinogenic risk depends on use of laboratory animal data and especially on several long-term bioassays. Syrian hamsters did not show a tumor response at any site at exposures up to 3,500 ppm for 6 hr/day 5 days/week, but mice and rats exposed at up to 4,000 ppm for 6 hr/day 5 days/week had treatment-related tumorigenic effects. EPA, after evaluating the data, classified methylene chloride as a probable human carcinogen (B2).

In accord with the default options of EPA's guidelines, the carcinogenic potency of methylene chloride was estimated by scaling the laboratory animal data to humans with a body surface-area conversion factor. The resulting cancer risk estimate was 4.1×10^{-6} for exposure at 1 $\mu g/m^3$ (Table 6-1). After further consideration, EPA has decreased this estimate by an order of magnitude (EPA, 1991d). The reduction is based on research on the pathways through which methylene chloride is metabolized. As with some other carcinogens, the risk of cancer arises not from methylene chloride itself, but rather from its metabolites. A correct calculation of the risk posed by methylene chloride therefore rests on understanding the human body's processes for metabolizing this chemical.

TABLE 6-1. Cancer incidence in B6C3F1 female mice exposed to methylene chloride and human cancer risk estimates derived from animal data

Animal Data

Concentration, Administered	Transformed Animal mg/kg, day	Human Equivalent mg/kg, day	Incidence of Liver Tumors	Incidence of Lung Tumors
4000	3162	712	40/46	41/46
2000	1582	356	16/46	16/46
0	0	0	3/45	3/45

Human Risk Estimates

Extrapolation Model	Cancer Risk[b] for 1 $\mu g/m^3$
LMS[a], surface area	4.1×10^{-6}
LMS, PB-PK[c]	3.7×10^{-8}
Logit	2.1×10^{-13}
Weibull	9.8×10^{-8}
Probit	$< 10^{-15}$
LMS-PB-PK with scaling for sensitivity	4.7×10^{-7}

[a] LMS = linearized multistage model.
[b] Upper 95% confidence limit.
[c] PB-PK = physiologically based pharmacokinetic.

Source: Modified from Reitz, et al., 1989.

Research with animal species used in the bioassays and human tissue has shed light on the metabolism of methylene chloride. Much of the research was conducted with the goal of providing input for physiologically based pharmacokinetic (PB-PK) models (Andersen et al., 1987, 1991; Reitz et al., 1989). The data were modeled in various ways, including consideration of two metabolic pathways. One involves oxidation by mixed-function oxidase (MFO) enzymes, and the other involves a glutathione-*S*-transferase (GST). Both pathways involve the formation of potentially reactive intermediates: formyl chloride in the MFO pathway and

chloromethyl glutathione in the GST-mediated pathway. The MFO pathway was modeled as having saturable, or Michaelis-Menten, kinetics, and the GST pathway as a first-order reaction, i.e., proportional to concentration. The analyses suggested that a reactive metabolite formed in the GST pathway was responsible for tumor formation. This pathway, according to the analyses, contributes importantly to the disposition of methylene chloride only at exposures that saturate the primary MFO pathway. The analyses further indicated that the GST pathway is less active in human tissues than in mice. This suggests that the default option of scaling for surface area yields a human risk estimate that is too high to be plausible. EPA incorporated the data on pharmacokinetics and metabolism into its most recent risk assessment for methylene chloride, although it retained a surface-area correction factor—now identifying it as a correction for interspecies differences in sensitivity. The new risk estimate is 4.7×10^{-7} for continuous exposure at 1 $\mu g/m^3$ (Table 6-1).

The process by which EPA arrived at the current risk estimate for methylene chloride with PB-PK modeling involved use of peer-review groups and SAB review to achieve a scientifically acceptable consensus position on the validity of the alternative model. After EPA's re-evaluation, however, articles in the peer-reviewed literature began to focus attention on parameter uncertainties in PBPK modeling, which neither EPA nor the original researchers in the methylene chloride case had considered. In the specific case of methylene chloride, at least one of the analyses (Portier and Kaplan, 1989) suggested that according to the new PBPK information EPA should have raised, rather than lowered, its original unit risk estimate if it wanted to continue to take a conservative stance. The more general point, which we discuss in Chapter 9, is that EPA must simultaneously consider both the evidence for departing from default models and the need to generate or modify the parameters that drive both the alternative and default models.

Formaldehyde

The toxicity and carcinogenicity of formaldehyde, a widely used commodity chemical, have been intensely studied and recently reviewed (Heck et al., 1990; EPA, 1991e). Concern for the potential human carcinogenicity of formaldehyde was heightened by the observation that exposure of rats at high concentrations (14.3 ppm) resulted in a very large increase in the incidence of nasal cancer. That observation gave impetus to the conduct and interpretation of epidemiologic studies of formaldehyde-exposed human populations. In the aggregate, the 28 studies that have been reported provide limited evidence of human carcinogenicity (EPA, 1991e). The "limited" classification is used primarily because the incidence of cancers of the upper respiratory tract has been confounded by exposure to other agents known to increase the rate of cancer, such as cigarette smoke and wood dusts.

The effects of chronic inhalation of formaldehyde have been investigated in rats, mice, hamsters, and monkeys. The principal evidence of carcinogenicity comes from studies in both sexes and two strains of rats and the males of one strain of mice, all showing squamous cell carcinomas of the nasal cavity.

The results of the rat bioassay have been used to derive quantitative risk estimates for cancer induction in humans (Kerns et al., 1983). Table 6-2 shows these animal data and the estimates of human cancer risk based on different exposure-dose models. (The table uses the inhalation cancer unit risk—the lifetime risk of developing cancer from continuous exposure at 1 ppm.) The 1987 EPA risk estimate (EPA, 1987c) measured exposure as the airborne concentration of formaldehyde. The rat bioassay shows a steep nonlinear exposure-response relationship for nasal-tumor induction. For example, two tumors were observed at 5.6 ppm, whereas 37 would have been expected from linear extrapolation from 14.3 ppm. Similarly, no tumors were observed at 2 ppm, whereas linear extrapolation from 14.3 ppm would have predicted 15.

TABLE 6-2. Incidence of nasal tumors in F344 rats exposed to formaldehyde and comparison of EPA estimates of human cancer risk associated with continuous exposure to formaldehyde

Exposure rate, ppm[a]	Incidence of Rat Nasal Tumors
14.3	94/140
5.6	2/153
2.0	0/159
0	0/156

Exposure Concentration, ppm	1987 Risk Estimates[b]	1991 Risk Estimates[c] Monkey-Based	1991 Risk Estimates[c] Rat-Based
	Upper 95% Confidence Limit Estimates		
1.0	2×10^{-2}	7×10^{-4}	1×10^{-2}
0.5	8×10^{-3}	2×10^{-4}	3×10^{-3}
0.1	2×10^{-3}	3×10^{-5}	3×10^{-4}
	Maximum Likelihood Estimates		
1.0	1×10^{-2}	1×10^{-4}	1×10^{-2}
0.5	5×10^{-4}	1×10^{-5}	1×10^{-3}
0.1	5×10^{-7}	4×10^{-7}	3×10^{-5}

[a] Exposed 6 hr/day, 5 days/week for 2 years.
[b] Estimated with 1987 inhalation cancer unit risk of 1.6×10^{-2} per ppm, which used airborne concentration as measure of exposure.
[c] Estimated with 1991 inhalation cancer unit risks of 2.8×10^{-3} per ppm (rat) and 3.3×10^{-4} per ppm (monkey), which used DNA-protein cross-links as measure of exposure.

Source: Adapted from EPA, 1991b.

The key issue became whether the same exposure-response relationship exists in people as in rats. To determine the answer, researchers directed substantial effort toward investigating the mechanisms by which formaldehyde exerted a carcinogenic effect. One avenue of investigation was directed toward characterizing DNA-protein cross-links as a measure of internal dose of formaldehyde (Heck et al., 1990). That work, initially conducted in rats, demonstrated a steep nonlinear relationship between formaldehyde concentration and formation of DNA-protein cross-links in nasal tissue, where most inhaled formaldehyde is deposited in rats. This suggested a correlation between such cross-links and tumors.

When the studies were extended to monkeys, a similar nonlinear relationship was observed between exposure concentration and DNA-protein cross-links in nasal tissue, but the concentration of DNA-protein cross-links per unit of exposure concentration was substantially lower than in the rat. Because the breathing patterns of humans more closely resemble those of monkeys than those of rats, the results of these studies suggested that using rats as a surrogate for humans might overestimate doses to humans, and hence the risk presented to humans by formaldehyde. EPA's most recent risk assessment (EPA, 1991e) used DNA-protein cross-links as the exposure indicator and estimated the human cancer risk (Table 6-2). EPA noted that the cross-links were being used only as a measure of delivered dose and that present knowledge was
insufficient to ascribe a mechanistic role to the DNA-protein cross-links in the carcinogenic process.

The EPA risk estimates for formaldehyde have been the subject of extensive peer review and review by the SAB. The 1992 update was reviewed by the SAB Environmental Health Committee and Executive Committee. The SAB recommended that the agency attempt to develop an additional risk estimate using the epidemiological data and prepare a revised document reporting all the risk estimates developed by the alternative approaches with their associated uncertainties. The two examples just discussed used mechanistic data and modeling to improve the characterization of the exposure-dose link. It is possible that as knowledge increases, models can be developed that link dose to response; the possibility is further discussed in Chapter 7.

The same is true of the linearized multistage model. As noted earlier, this model assumes that risk is linear in dose. As noted earlier, however, rats exposed to formaldehyde show a steep nonlinear exposure-response relationship. This raises the possibility that the linearized multistage model might be inappropriate for at least some chemicals. It is possible that advances in knowledge of the molecular and cellular mechanisms of carcinogenesis will show a need to use other models either case by case or generically. More discussion of this matter can be found in Chapter 7.

The strategy advocated for formaldehyde would build on multistage models of the carcinogenic process that describe the accumulation of procarcinogenic mutations in target cells and the consequent malignant conversion of these cells (Figure 6-1). The Moolgavkar-Venzon-Knudson model substantially oversimplifies the carcinogenic process but provides structural framework for integrating and examining data on the role of DNA-protein cross-links, cell replication, and other biologic phenomena in formaldehyde-induced carcinogenesis (Moolgav-

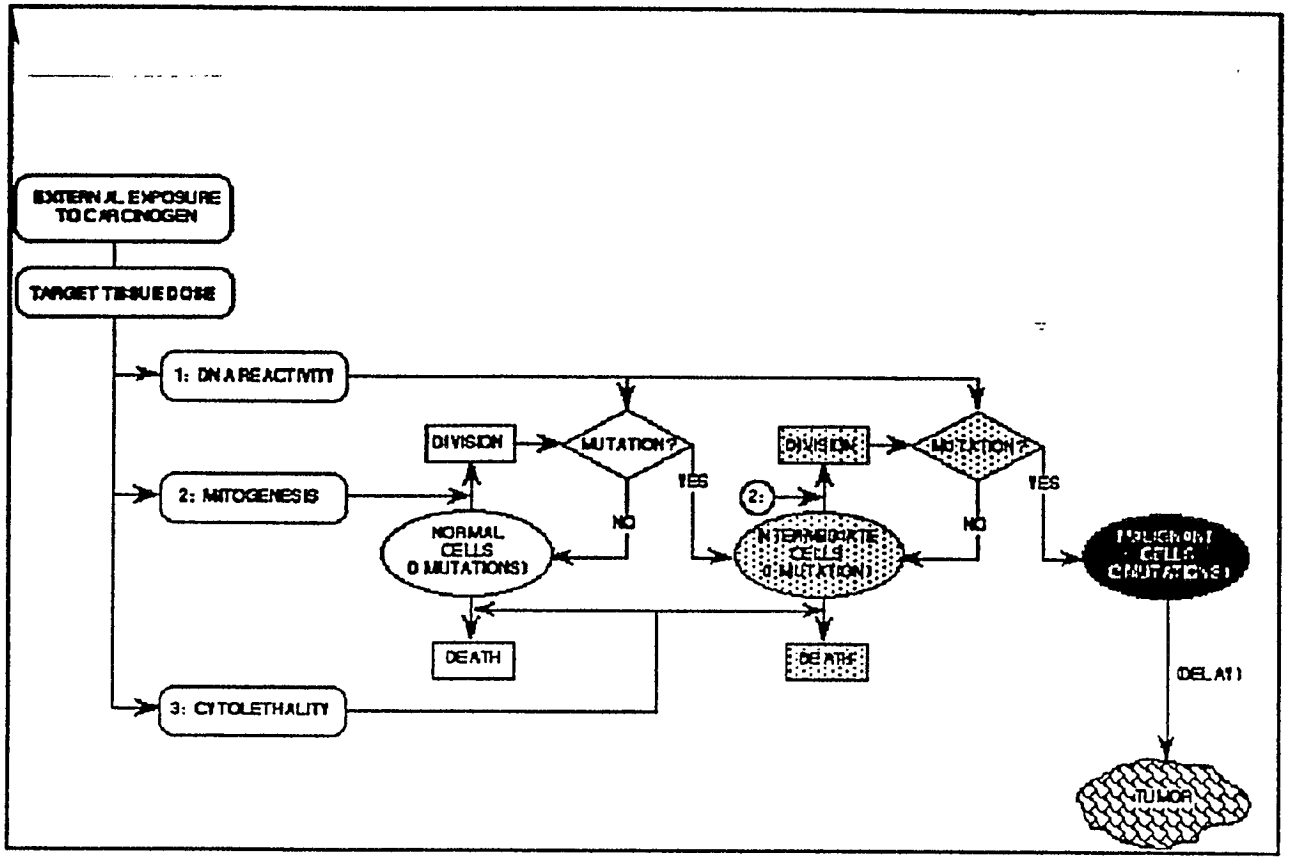


Figure 6-1 Model of chemical carcinogenesis built on multi-stage carcinogenesis model of Moolgavkar-Vinson-Knudson (Conolly et al., 1992)

kar and Venzon, 1979; Moolgavkar and Knudson, 1981; Moolgavkar et al., 1988; NRC, 1993b). Key features of this model are definition of the relationship of target-tissue dose to exposure and the use of that dose as a determinant of three outcomes: reactivity with DNA, mitogenic alterations, and cytolethality. These, in turn, cause further biologic effects: DNA reactivity leads to mutations, the mitogenic stimuli increase the rate of cell division, and cells die (cell death stimulates compensatory cell proliferation). Models like that shown provide a structured approach for integrating data on a toxicant, such as formaldehyde. It is anticipated that modeling will provide insight into the relative importance, at various exposure concentrations, of the two mechanisms that appear to have a dominant role in formaldehyde carcinogenesis: mutation and cell proliferation. Improved insight into their role could provide a mechanistic basis for selecting between the linearized multistage mathematical model now used for extrapolation from high to low doses and alternative models that might have more biologic plausibility.

TRICHLOROETHYLENE

Trichloroethylene (TCE) is a chlorinated solvent that has been widely used in the industrial degreasing of metals. TCE is a concern to EPA as an air pollutant, a water pollutant, and a substance frequently present in ground water at Superfund sites. EPA carried out a risk assessment for TCE documented in a health assessment document (HAD) (EPA, 1985d) and a draft addendum incorporating additional inhalation-bioassay data (EPA, 1987e). Both documents were reviewed by the SAB (EPA, 1984a; EPA, 1988j,k). The second document has not been issued in final form, and no further revision of EPA's risk assessment on TCE has been made since 1987.

The carcinogenic potency of TCE is based on the liver-tumor response in B6C3F1 mice, a strain particularly prone to liver tumors. The carcinogenicity of TCE might result from trichloroacetic acid (TCA), a metabolite of TCE that is itself known to cause liver tumors in mice. TCA is one of a number of chemicals that cause proliferation of peroxisomes, an intracellular organelle, in liver cells. Peroxisome proliferation has been proposed as a causal mechanism for the liver tumors, and proponents have asserted that such tumors should receive treatment in risk assessments different from evaluation under EPA's default assumptions. In particular, human liver cells might be much less sensitive than mouse liver cells to tumor formation from this mechanism, and the dose-response relationship might be nonlinear at low doses.

The SAB held a workshop in 1987 on peroxisome proliferation as part of its reviews on risk assessments for TCE and other chlorinated solvents. While endorsing a departure from the default on the alpha-2μ-globulin mechanism described in example 1 above, the SAB declined to endorse such a departure for peroxisome proliferation, noting that a causal relationship for this mechanism was "plausible but unproven." The SAB strongly encouraged further research, describing this mechanism for mouse liver tumors as "most promising for immediate application to risk assessment" (EPA, 1988k). The SAB criticized EPA on the draft Addendum on

TCE (EPA, 1987e) for not adequately presenting uncertainties and for not seriously evaluating recent studies on the role of peroxisome proliferation (EPA, 1988*l*).

In the TCE case, departure from the defaults was rejected after an SAB review that recognized the peroxisome proliferation mechanism as plausible. Controversy over the interpretation of liver tumors in B6C3F1 mice continues. Some scientists assert that EPA's use of the tumor-response data from this particularly sensitive strain has been inappropriate (Abelson, 1993; ILSI, 1992). In the TCE example, departure from the defaults might become appropriate, on the basis of improved understanding of mouse liver tumors and their implications for human cancer. Although the SAB declined to endorse such a departure in 1987, it strongly encouraged further research as appropriate for supporting improved risk assessment.

CADMIUM

Cadmium compounds are naturally present at trace levels in most environmental media, including air, water, soil, and food. Substantial additional amounts might result from human activities, including mining, electroplating, and disposal of municipal wastes. EPA produced an HAD on cadmium (EPA, 1981b) and later an updated mutagenicity and carcinogenicity assessment (EPA, 1985e). The latter went through SAB review (EPA, 1984b), which pointed out many weaknesses and research needs for improving the risk assessment. No revision of the risk assessment on cadmium has occurred since 1985.

EPA used epidemiological data for developing a single unit risk estimate for all cadmium compounds. Use of the estimate from the best available bioassay would have given a unit risk for cadmium compounds higher by a factor of 50. The SAB and EPA in its response to SAB comments (EPA, 1985f) agreed that the solubility and bioavailability of different cadmium compounds were important in determining the risk associated with different cadmium compounds and that such differences might explain the discrepancy between the epidemiological data and the bioassay data. No implementation of the principle that cadmium compounds should be evaluated on the basis of bioavailability has yet been devised, although its importance to risk assessment for some air pollutants that contain cadmium is clearly set forth in EPA's response to the SAB (EPA, 1985f).

EPA's existing risk assessment for cadmium might be judged adequate for screening purposes. But the SAB review and the EPA response to it suggest that the carcinogenic risk associated with a specific cadmium compound could be overestimated or underestimated, because bioavailability has not been included in the risk assessment. A refined version of the risk assessment that includes bioavailability might be appropriate, especially if residual risks for cadmium compounds appear to be important under the Clean Air Act Amendments of 1990.

NICKEL

Nickel compounds are found at detectable levels in air, water, food, and soil. Increased concentrations of airborne nickel result from mining and smelting and from combustion of fuel that contains nickel as a trace element. Nickel compounds present in smelters that use the pyrometallurgical refining process are clearly implicated as human carcinogens. EPA's HAD on nickel (EPA, 1986b) lists dust from such refineries and nickel subsulfide as category A (known human) carcinogens. A rare nickel compound, nickel carbonyl, is listed, on the basis of sufficient evidence in animals, as category B2. Other nickel compounds are not listed as carcinogens, although EPA states (EPA, 1986b, p. 2-11) :

> The carcinogenic potential of other nickel compounds remains an important area for further investigation. Some biochemical and *in vitro* toxicological studies seem to indicate the nickel ion as a potentially carcinogenic form of nickel and nickel compounds. If this is true, all nickel compounds might be potentially carcinogenic with potency differences related to their ability to enter and to make the carcinogenic form of nickel available to a susceptible cell. However, at the present time, neither the bioavailability nor the carcinogenesis mechanism of nickel compounds is well understood.

The SAB reviewed the nickel HAD and concurred with EPA's listing of only the three rare nickel species as category A and B2 carcinogens (EPA, 1986c).

The results of bioassays on three nickel species by the National Toxicology Program are due to be released soon, and these results should provide a basis for revision of risk assessments for nickel compounds.

The cadmium and nickel examples point out an important additional default option: Which compounds should be listed as carcinogens when it is suspected that a class of chemical compounds is carcinogenic? Neither the cadmium risk assessment, the nickel risk assessment, or EPA's *Guidelines for Carcinogen Risk Assessment* (EPA, 1986a) provide specific guidance on this issue.

DIOXINS

Dioxins are a class of organochlorine compounds that can form as the result of the combustion or synthesis of hydrocarbons and chlorine-containing substances. One isomer, 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD), is one of the most potent carcinogens ever tested in bioassays. EPA issued an HAD for dioxins (EPA, 1985g), which the SAB criticized for its treatment of the non-TCDD isomers that may contribute substantially to the overall toxicity of a mixture of dioxins (EPA, 1985h).

The potency calculation for TCDD has continued to be a subject of controversy. Research indicates that the toxic effects of TCDD may result from the binding of TCDD to the Ah (aromatic hydrocarbon) receptor. In 1988, EPA asked the SAB to review a proposal to revise its risk estimate for TCDD. SAB agreed with EPA's criticism of the linearized multistage model and its assessment of the promise of alternative models based on the receptor mecha-

nism. But SAB did not agree that there was adequate scientific support for a change in the risk estimate. SAB carefully distinguished its recommendation from a change that EPA might wish to make as part of risk management (EPA, 1989f)

> The Panel thus concluded that at the present time the important new scientific information about 2,3,7,8-TCDD does not compel a change in the current assessment of the carcinogenic risk of 2,3,7,8-TCDD to humans. EPA may for policy reasons set a different risk-specific dose number for the cancer risk of 2,3,7,8-TCDD, but the Panel finds no scientific basis for such a change at this time. The Panel does not exclude the possibility that the actual risks of dioxin-induced cancer may be less than or greater than those currently estimated using a linear extrapolation approach.

A recent conference affirmed the scientific consensus on the receptor mechanism for TCDD, but there was not a consensus that this mechanism implied a basis for departure from low-dose linearity (Roberts, 1991). After the conference, and after the recommendations of the SAB (EPA, 1989f), EPA initiated a new study to reassess the risk for TCDD. That study is now in draft form and scheduled for SAB review in 1994.

The potencies of other dioxin isomers and isomers of a closely related chemical class, dibenzofurans, have been estimated by EPA with a toxic-equivalency-factor (TEF) method (EPA, 1986d). The TEF method was endorsed by the SAB as a reasonable *interim* approach in the absence of data on these other isomers (EPA, 1986e). The SAB urged additional research to collect such data. Municipal incinerator fly ash was used as an example of a mixture of isomers of regulatory importance that might be appropriate for long-term animal testing.

The EPA initiative for a review of TCDD is one of the few instances in which the agency has initiated revision of a carcinogen risk assessment on the basis of new scientific information. Dioxins and dibenzofurans are unique in that potency differences within this class of closely related chemical isomers are dealt with through a formal method that has undergone peer review by the SAB.

EXAMPLE 3: MODELING EXPOSURE-RESPONSE RELATIONSHIP

If chemicals act like radiation at low exposures (doses) inducing cancer—i.e., if intake of even one molecule of a chemical has an associated probability for cancer induction that can be calculated—the appropriate model for relating exposure-response relationships is a linearized multistage model.

Of the 189 hazardous air pollutants, unit risk estimates are available for only 51: 38 with inhalation unit risks, which are applicable to airborne materials, and 13 with oral unit risks. The latter probably have less applicability to estimating the health risks associated with airborne materials. All 38 inhalation unit risk values have been derived with a linearized multistage model; i.e., it is assumed that the chemicals act like radiation. That might be an appropriate assumption for chemicals known to affect DNA directly in a manner analogous to that of radiation. For other chemicals—e.g., such nongenotoxic chemicals as chloroform—the

assumption of a mode of action similar to that of radiation might be erroneous, and it would be appropriate to consider the use of biologically-based exposure-response models other than the linearized multistage model.

The process of choosing between alternative exposure-response models is difficult because the models cannot be validated directly for their applicability for estimating lifetime cancer risks at exposures of regulatory concern. Indeed, it is possible to obtain cancer incidence data on exposed laboratory animals and distinguish them from the control incidence only over a narrow range, from some value over 1% (10^{-2}) to about 50% (5×10^{-1}) cancer incidence. In regulation of chemicals, the extrapolation may be over a range of up to 4 orders of magnitude (from 10^{-2} to 10^{-6}), going from experimental observations to estimated risks of cancer incidence at exposures of regulatory concern. One approach to increasing the accuracy with which comparisons between measured outcome and model projections can be made involves increasing the size of the experimental populations. However, statistical considerations, the cost of studying large numbers of animals, and the greater difficulty of experimental control in larger studies put narrow limitations on the use of this approach. Similar problems exist in conducting epidemiological studies.

An attractive alternative is to use advances in knowledge of the molecular and cellular mechanisms of carcinogenesis. Identification of events (e.g., cell proliferation) and markers (e.g., DNA adducts, suppressor genes, oncogenes, and gene products) associated with various steps in the multistep process of carcinogenesis creates a potential for modeling these events and products at low exposure. Direct tests of the validity of exposure-response models at risks of around 10^{-6} are not likely in the near future. However, with an order-of-magnitude improvement in sensitivity of detection of precancerous events with a probability of occurrence down to around 10^{-3}-10^{-2}, the opportunity will be available to evaluate alternative modes of action and related exposure-response models at substantially lower exposure concentrations than has been possible in the past. For example, it should soon be possible to evaluate compounds that are presumed to have different modes of action (direct interaction with DNA and genotoxicity versus cytotoxicity) and alternative models (linearized multistage versus nonthreshold) that might yield markedly different risks when extrapolated to realistic exposures and low risks.

FINDINGS AND RECOMMENDATIONS

USE OF DEFAULT OPTIONS

FINDING: EPA's practice of using default options when there is doubt about the choice of appropriate models or theory is reasonable. EPA should have a means of filling the gap when scientific theory is not sufficiently advanced to ascertain the correct answer, e.g., in extrapolating from animal data to responses in humans.

RECOMMENDATION: EPA should continue to regard the use of default options as a reasonable way to cope with uncertainty about the choice of appropriate models or theory.

Articulation of Defaults

FINDING: EPA does not clearly articulate in its risk-assessment guidelines that a specific assumption is a default option.

RECOMMENDATION: EPA should clearly identify each use of a default option in future guidelines.

Justification for Defaults

FINDING: EPA does not fully explain in its guidelines the basis for each default option.

RECOMMENDATION: EPA should clearly state the scientific and policy basis for each default option.

Alternatives to Default Options

FINDING: EPA's practice appears to be to allow departure from a default option in a specific case when it ascertains that there is a consensus among knowledgeable scientists that the available scientific evidence justifies departure from the default option. EPA, though, has not articulated criteria for allowing departures.

RECOMMENDATION: The agency should consider attempting to give greater formality to its criteria for a departure, to give greater guidance to the public and to lessen the possibility of ad hoc, undocumented departures from default options that would undercut the scientific credibility of the agency's risk assessments. At the same time, the agency should be aware of the undesirability of having its guidelines evolve into inflexible rules.

Process for Departures

FINDING: EPA has relied on its Science Advisory Board and other expert bodies to determine when a consensus among knowledgeable scientists exists.

RECOMMENDATION: EPA should continue to use the Science Advisory Board and other expert bodies. In particular, the agency should continue to make the greatest possible use of

peer review, workshops, and other devices to ensure broad peer and scientific participation to guarantee that its risk-assessment decisions will have access to the best science available through a process that allows full public discussion and peer participation by the scientific community,

MISSING DEFAULTS

FINDING: EPA has not stated all the default options in each step in the risk-assessment process, nor the steps used when there is no default. Chapters 7 and 10 elaborate on this matter and identify several possible "missing defaults."

RECOMMENDATION: EPA should explicitly identify each generic default option in the risk-assessment process.

7
MODELS, METHODS, AND DATA

INTRODUCTION

Health risk assessment is a multifaceted process that relies on an assortment of methods, data, and models. The overall accuracy of a risk assessment hinges on the validity of the various methods and models chosen, which in turn are governed by the scope and quality of data. The degree of confidence that one can place in a risk assessment depends on the reliability of the models chosen and their input parameters (i.e., variables) and on how well the boundaries of uncertainty have been quantified for the input parameters, for the models as a whole, and for the entire risk-assessment process.

Quantitative assessment of data quality, verification of method, and validation of model performance are paramount for securing confidence in their use in risk assessment. Before a data base is used, the validity of its use must be established for its intended application. Such validation generally encompasses both the characterization and documentation of data quality and the procedures used to develop the data. Some characteristics of data quality are overall robustness, the scope of coverage, spatial and temporal representativeness, and the quality-control and quality-assurance protocols implemented during data collection. More specific considerations include the definition and display of the accuracy and precision of measurements, the treatment of missing information, and the identification and analysis of outliers. Those and similar issues are critical in delineating the scope and limitations of a data set for an intended application.

The performance of methods and models, like that of data bases, must be characterized and verified to establish their credibility. Evaluation and validation procedures for a model might include sensitivity testing to identify the parameters having the greatest influence on the output values and assessment of its accuracy, precision, and predictive power. Validation of a model also requires an appropriate data base.

This chapter discusses the evaluation and validation of data and models used in risk assessment. In cases where there has been an insufficient assessment of performance or quality, research recommendations are made. Although in this chapter we consider validation issues sequentially, according to each of the stages in the (modified) Red Book paradigm, our goal here is to make the assessment of data and model quality an iterative, interactive component of the entire risk-assessment and risk-characterization process.

EMISSION CHARACTERIZATION

As described in Chapter 3, emissions are characterized on the basis of emission factors, material balance, engineering calculations, established Environmental Protection Agency (EPA) protocols, and measurement. In each case, this characterization takes the structural forms of a linearly additive process (i.e., emissions equals product - [feedstock + accumulations]), a multiplicative model (i.e., emissions equals [emission factor][process rate]), or an exponential relationship (e.g., emission equals intercept + [(emission factor) (measurement)^{exp}]).

The additive form is based on the mass-balance concept. An estimate is made by measuring the feedstock and product to determine an equipment-specific or process-specific transfer coefficient. This coefficient is used to estimate emissions to the atmosphere. The measurements available for the additive form are often not sufficiently precise and accurate to yield complete information on inputs and outputs (NRC, 1990a). For example, an NRC committee (NRC, 1990a) considered a plant that produced 5 million pounds of ethylene per day and used more than 200 monitoring points to report production with a measurement accuracy of 1%, equivalent to 50,000 lb of ethylene per day. The uncertainty in this estimate (50,000 lb) greatly exceeded a separate estimate of emissions, 191 lb, which was calculated by the plant and was confirmed by monitoring of the emission points. Thus, despite the apparently good precision of estimates within 1%, the additive method was not reliable. This seems to be generally true for complicated processes or multiple processing steps.

The other forms are based on exponential and multiplicative models. Each may be deterministic or stochastic. For example, emissions from a well-defined sample of similar sources may be tested to develop an emission factor that is meant to be representative of the whole population of sources. A general difficulty with such fits that use these functional (linear or one of several nonlinear forms) forms is that the choice of form may be critical but hard to validate. In addition, it must be assumed that data from the sources used in the calculations are directly applicable to the sources tested in process design and in the management and maintenance approaches of the organizations that run them are the same in all cases.

An example of an exponential form of an emission calculation is shown in Figure 7-1. This figure shows the correlation between screening value (the measurement) and leak rate (the emission rate) for fugitive emissions from a valve. The screening value is determined by measuring the hydrocarbons emitted by a piece of equipment (in this case, a valve in gas service) with an instrument like an OVA (organic-vapor analyzer). The leak rate (i.e., emission) is then determined by reading the value on the y axis corresponding to that screening value. Note that the plot is on a log-log scale, so that a "3" on the x axis indicates that a 1,000-ppm screening value corresponds to a "-3.4" on the y axis, or 0.001 lb/hr for each valve in gas service at that screening value. The observations here are based on an analysis conducted for 24 synthetic organic chemical manufacturing industry (SOCMI) units representing a cross-section of this industry (EPA, 1981a).

As part of this analysis, a six-unit maintenance study (EPA, 1981a) was used to determine the impact of equipment monitoring and maintenance using an OVA instrument on emission

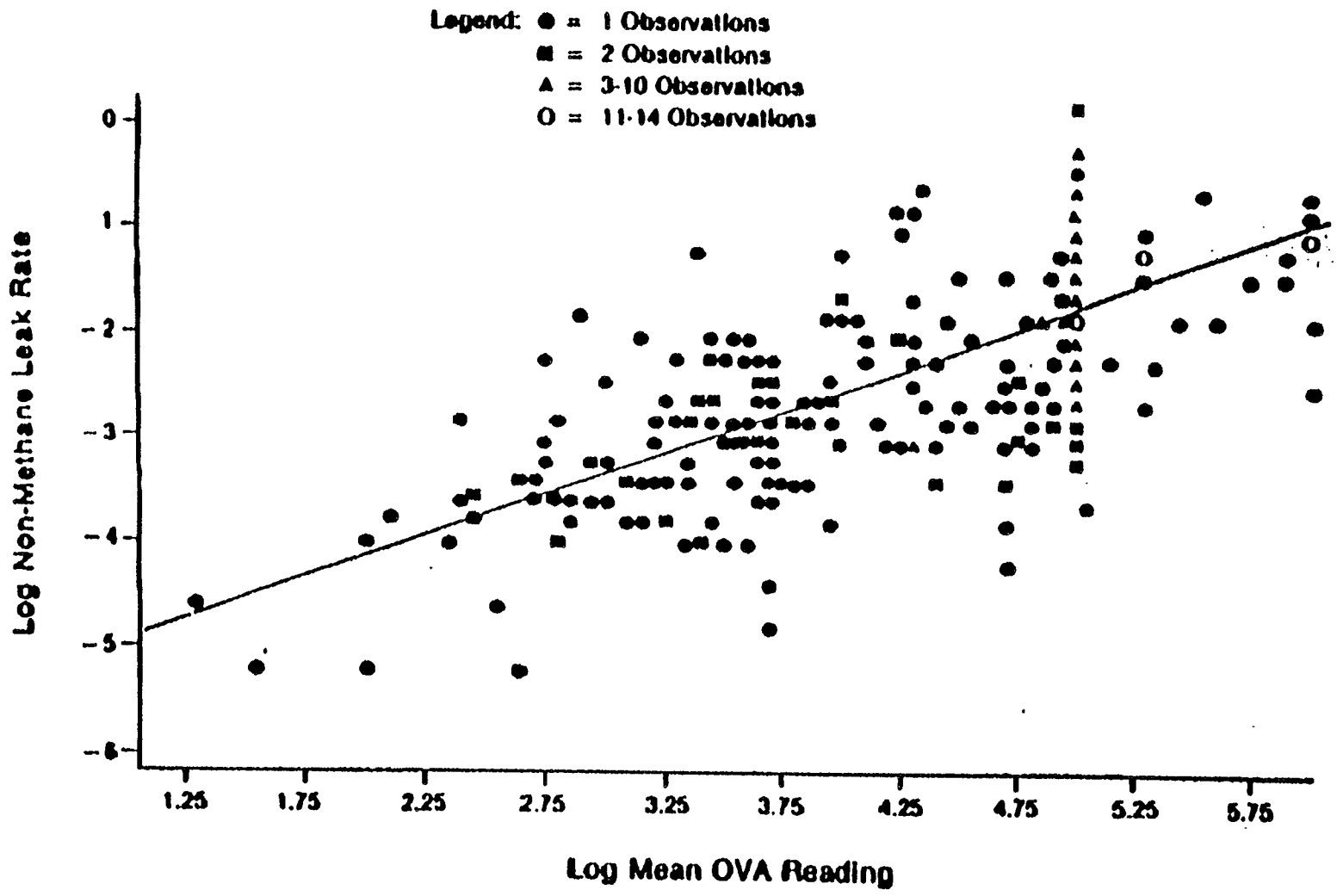


Figure 7-1. Log_{10} leak rate vs. log_{10} OVA reading for values-gas service. Source: EPA, 1981a.

reduction. The equation derived for the valve emissions in gas service explains only 44% (square of the correlation coefficient) of the variance in the points shown in Figure 7-1. Similar results were obtained from other possible emission points.

The facilities in this SOCMI study could reduce the estimate of their emissions by 29-99% by determining plant-specific emission factors, indicating the difficulties in using industry-wide average to represent specific plant behavior.

The multiplicative form improves on the emission-factor approach, in that it incorporates more features of the process, attempting to accommodate the types of equipment being used, the physical properties of the chemical, and the activity of the equipment as a whole. The deterministic form of the multiplicative model is based on the chemical and physical laws that determine the emission rate. The variables measured—vapor pressure, molecular weight, temperature, etc.—are chemical physical properties that are related to the emission rate. The multiplicative form provides some scientific basis for the estimate beyond the simple curve-fitting. However, it has difficulties, because some of the properties are not constant. For example, the ambient air temperature, one factor in determining the emission rate, can vary quite widely within a day. The average temperature for a given period, such as a month, is used for ease in calculation, but this practice introduces some error. EPA might want to consider a more detailed analysis in which the emissions that occur during the period are stratified into groups with smaller variations in variables such as ambient temperature. The emissions in the strata could be estimated and weighted sums calculated to provide a better estimate.

Probably the most accurate procedure is to use none of those "forms" to determine emissions, but rather to sample stack and vent emissions at each source. However, such sampling can be quite expensive, and the costs could overburden owners of small sources. Apart from costs, the primary difficulty with this procedure is that it yields an estimate for one site on one occasion. Emissions could change because of a variety of factors. An alternative to testing is to estimate emissions from monitoring data. Continuous emission monitors (CEMs), which are available for a small number of chemicals, are placed in stacks or near fugitive-emission points to measure the concentration of a chemical being released; concentrations can then be converted to amounts. However, CEMs can be expensive and difficult to maintain, and they may produce incomplete or inaccurate measurements. When such testing is conducted, however, they may show that other kinds of estimates are seriously in error. For example, a study (Amoco/EPA, 1992) compared emissions estimated primarily from emission factors with those determined during testing. The measured overall actual estimate of emissions was more than twice as high as the TRI estimate for a variety of reasons, including identification of new sources, overestimation or underestimation of the importance of some sources, and the lack of a requirement to report source emissions under a particular regulation.

EVALUATION OF EPA PRACTICE

EPA has worked diligently to help members of the public who are required to provide

emission estimates for regulatory purposes. This 20-year effort has provided documents that are used to estimate air-pollutant emissions throughout the world. However, in some cases, EPA has had to provide emission estimation factors based on very little information about the process involved; it was difficult to check the assumption that the process for which the calculation is being used is similar to the process that was tested in the development of the emission factor.

There are two basic difficulties with the way EPA applies its emission estimation techniques. First, most estimates are made by using the emission factors or by fitting the linear or exponential forms. As discussed previously, the accuracy of emission estimates using these techniques might not be high.

Second, the information is generated in such a way that only point estimates are presented. Although it is clear from the earlier discussion that there can be uncertainty in the estimates, EPA has extensive files on how the emission factors were determined, and this information presumably contains enough points to generate distribution of emissions rather than just a point estimate. EPA provides only qualitative ratings of the accuracy of the emission method. The ratings are not based on the variance in the estimate, but just on the number of emission points used to generate the data. If there are enough points to generate an emission factor, it is possible to estimate the distribution of emission factors from which an estimate can be chosen to solve a particular exposure-risk estimation problem.

However, the emission factors are given only a "grade" from A (best) to E relative to the quality and amount of data on which estimates are based. An emission factor based on 10 or more plants would likely get an "A" grade, whereas a factor based on a single observation of questionable quality or one extrapolated from another factor for a similar process would probably get a D or E. The grades are subjective and do not consider the variance in the data used to calculate factors. According to EPA (1988e), the grades should "be used only as approximations, to infer error bounds or confidence intervals about each emission factor. At most, a [grade] should be considered an indicator of the accuracy and precision of a given factor used to estimate emissions from a large number of sources." The uncertainty in the estimates is such that EPA is not comfortable with the A-E system and is developing a new qualitative system to indicate uncertainty. EPA is attempting to generate estimation factors for hazardous air pollutants industry by industry, but it is still hesitant to ascribe any sort of uncertainty to emission factors.

A single disruption in operation of a plant can increase the release rate for some interval (hour or day). An extreme example is the dioxin release from a manufacturing plant in Seveso, Italy. Such disruptions are not incorporated into any of the emission characterizations, except for the few cases where emission monitoring is available. However, in those cases, emissions might be so high that they exceed the maximum reading of a monitor and thereby lead to just a lower bound (if this problem is recognized) or even to a serious underestimate of the actual emission. Furthermore, the frequency and duration of such episodes are unpredictable.

Therefore, EPA should also attempt to make some sort of quantitative estimates of the variability of measured emissions among sources within a category and of the uncertainty in its

overall emission estimates for individual sources and the source category as a whole. This issue is discussed in more depth in Chapter 10, but could involve analyzing the four kinds of circumstances as appropriate for a particular source type—routine, regular maintenance, upsets and breakdowns, and rare catastrophic failures. EPA could also note the implications of the dynamics of causation of different effects for emission estimation, and the resulting need for estimates of exposure and exposure variability over different averaging times.

The itemization of emissions by chemical constituent also raises problems. Emission characterization methods often provide only the amount of VOCs (volatile organic compounds) that is emitted. The amounts of particular compounds (benzene, toluene, xylene, etc.,) within these VOC emissions are often not individually reported. Without the emission data on particular compounds, it is impossible to provide the information needed for exposure modeling in the risk-assessment process.

EPA does not appear to be making major strides toward improving the methods used to evaluate emissions. Although EPA is making extensive efforts to distribute the emission factors it has generated, the committee has found insufficient effort either to evaluate the accuracy of the underlying method used to derive the emission estimates or to portray the uncertainty in the emission factors. The primary exception is a joint effort of the Chemical Manufacturers Association (CMA) and EPA on fugitive emissions called Plant Organization Software System for Emission Estimation or POSSEE (CMA, 1989). In this case, companies are testing fugitive emissions within plants and collecting data on chemical and physical variables to derive emission estimates based on deterministic models (which use physical and chemical properties), rather than stochastic models. There have been efforts to increase the scientific justification of estimates of emissions from storage tanks: the American Petroleum Institute has developed data that have been used for developing the estimation method shown in the multiplicative form described above. The question then arises as to how to approach emission estimates in exposure assessments and risk assessments. The uncertainty in the mass-balance approach (additive form) can be so large that its use should be discouraged for any purposes other than for a very general screening. It is unlikely that an emission estimate derived with this method would be appropriate for risk assessment.

The linear emission-factor approach could be used as a general screening tool in an exposure assessment. As indicated by EPA in response to a question from this committee:

> While emission factor-based estimates can be useful in providing a general picture of emissions across an entire industrial category, use of such factors to provide inputs to a site-specific risk assessment may introduce a great deal of uncertainty into that assessment.

If such an approach is used for an entire industrial category, then at least the uncertainty of each emission factor should be determined. If there is enough information to derive an emission factor, then a probability distribution could be calculated. There may then be disagreement about where on the probability distribution the emission estimate should be chosen. However, it is better to make the choice explicitly, as discussed in Chapter 9. The

same situation is true for emissions estimated with the exponential and multiplicative approaches. EPA should include a probability distribution in all its emission estimates.

One method to determine the uncertainty in an emission estimate more easily would be to require each person submitting an emission estimate (for SARA 313 requirements, permitting, etc.) to include an evaluation of the uncertainty in the estimate. EPA could then evaluate the uncertainty in the estimation methods to determine whether the estimation was done properly. Although that might increase the costs of developing submissions slightly, the organization submitting the estimate might benefit from the results. Small sources unable to afford such analysis could instead define a range that is consistent with known or readily determined factors in their operation (e.g., for a dry cleaner, the pounds of clothes per week and gallons of solvent purchased each month).

EPA is reviewing, revising, and developing emission estimation methods for sources of the 189 chemicals. It is focusing on adding data, rather than evaluating its basic approach—the use of a descriptive model, instead of a model based on processes, for emission estimation. It appears from the examples given above that the uncertainties in emissions can dominate an exposure assessment and that a concerted effort to improve emission estimation could serve to substantially reduce the uncertainty in many risk estimates. Combined industry efforts to improve the techniques used to estimate fugitive emissions on the basis of physical and chemical properties (not just curve-fitting) should be encouraged.

EXPOSURE ASSESSMENT

Once an emission characterization is developed, it becomes one of the inputs into an air-quality model to determine the amount of a pollutant in ambient air at a given location. A population-exposure model is then used to determine how much of a pollutant reaches people at that location.

POPULATION

The size of the population that might be exposed to an emission must be determined. Population data have been collected, published, and scrutinized for centuries. Many such data refer to entire populations or subpopulations, so questions of representation and statistical aspects of sampling do not arise in their usual form. Even where sampling is used, a large background of technique and experience allows complex estimation and other kinds of modeling to proceed without the large uncertainties inherent in, for example, extrapolation from high to low doses of toxic agents or from rodents to humans.

Population data are almost always affected to some degree by nonsampling error (bias), but this is well categorized, understood, and not a serious problem in the context of risk assessment. For example, terminal-digit preference (e.g., a tendency to report ages that end in zero or five) has been minimal since the attainment of nearly universal literacy and especially since

the adoption of birth certification. Attainment of advanced ages (i.e., over 80 years) is still overstated, but this is not quantitatively serious in age estimation for purposes of risk assessment (because EPA still assumes that 70 years is the upper-bound value of the length of a liftime). Population undercounts in the U.S. census of 1990 averaged about 2.1% and were substantially higher for some subgroups, perhaps up to 30%; however, even 30% uncertainty is smaller than many other sources of error that are encountered in risk assessment. The largest proportionate claim of uncertainty seems to be in the number of homeless persons in the United States; estimated uncertainty is less than a factor of 10.

Estimation of characteristics in groups or subgroups not examined directly is subject to additional uncertainty. For example, the 1992 population is not directly counted, but standard techniques are used to extrapolate from the census of 1990, which was a nearly complete counting of the population. Investigators have found earlier years estimates to be generally quite accurate, whether the extrapolations were strictly mathematical (e.g., based on linear extrapolation) or demographic (based on accounting for the addition of 3 years between 1990 and 1993, with adjustments for deaths, for births of the population under age 3, deaths, and net migration). The problems are greater for states and smaller areas, because data on migration (including internal migration) are not generally available.

Error tends to increase as subgroups get smaller, partly because statistical variability increases (i.e., small sample size leads to less precision in the estimate of the central tendency with any distributed measurement), but also because individual small segments are not as well characterized and as well understood as larger aggregates and because population data are generally collected according to a single nationwide protocol that allows for little deviation to accommodate special problems.

The committee is comfortable about using published population data for nearly all population characteristics and subgroups. Where adjustment to reduce errors is feasible, it should be used; but in the overall context of risk assessment, error in population assessment contributes little to uncertainty.

In some cases, a research study must define and identify its own population without help from official census and surveys. An example is a long-term followup study of workers employed in a specific manufacturing plant. When such studies are done by skilled epidemiologists, total counts, ages, and other demographic items tend to be accurate to within a factor of 2 or 3. The largest uncertainties are likely to be in the estimation of exposure to some toxic agent; these are often dealt with by the use of rough categories (high, medium, and low exposure) or surrogate measures (e.g., years employed in a plant, rather than magnitude of exposure). Errors in such work are of great concern, but they tend to be peculiar to each study and hence lead to study-specific remedies in design, performance, or analysis. They tend to be smaller than other kinds of uncertainties, but can still be of concern if a putative effect is also small.

As indicated, population data derived from a census and fortified with estimation methods are regarded as accurate and valid, and uncertainties introduced into risk assessment are relatively small. There is a need, however, for information on additional population characteristics that are not included in the census. There is a paucity of activity-pattern information,

and population-exposure models or individual-exposure-personal-exposure models have not been adequately tested or validated, because they use people's activity to estimate exposure to chemicals in air. Only a few small efforts have been undertaken to develop such a data base, namely, EPA's Total Exposure and Assessment Methodology (TEAM) program and the California EPA's State Activity Pattern Study. Those programs have acquired information about people's activities that cause the emission of air pollutants or place people in micro-environments containing air pollutants that potentially lead to exposure. There is a need to develop a national data base on activity patterns that can be used to validate models that estimate personal exposure to airborne toxic chemicals. Accurately described activity patterns coupled with demographic characteristics (e.g., socioeconomic) can be used for making a risk assessment and assessing the environmental equity of risk across socioeconomic groups and race.

When exposure-characterization models are developed for use in risk assessment, the bias and uncertainty that they yield in the calculation of exposure estimates should be clearly defined and stated, regardless of whether activity patterns are included. Later, the choice of an appropriate model from an array of possibilities should be based on, but not necessarily limited to, its quantitative measure of performance and its rationale should be included with a statement of the criteria for its selection.

Air-Quality Model Evaluation

Air-quality models are powerful tools for relating pollutant emissions to ambient air quality. Most air-quality models used in assessing exposure to toxic air pollutants have been extensively evaluated with specific data sets, and their underlying mathematical formulations have been critically reviewed. Relative to some of the other models for risk assessment of air pollutants, air-quality models probably enjoy the longest history of model evaluation, refinement, and re-evaluation. For example, the original Gaussian-plume models were formulated and tested in the 1950s. That does not mean, however, that model evaluation does not still continue or that the model evaluation should be dismissed in assessing air-pollutant exposure; in fact, previous studies have shown the benefits of model evaluation in every application.

Evaluation of the air-quality models and other components of air-pollutant risk assessment is intended to determine accuracy for providing the details required in a given application and to provide confidence in the results. In air-quality modeling, that is particularly important. A Gaussian-plume model, when used with the input data generally available, might not correctly predict where maximal concentrations will be realized (e.g., because winds at the nearest station, such as an airport, might differ in direction from winds near the source of interest), but should provide a reasonable estimate of the distribution of pollutant concentrations around the site. That might be sufficient for some applications, but not others. Model evaluation can also add insight as to whether a tool is "conservative" or the opposite, and it can provide a quantitative estimate of uncertainty.

Of particular concern are the more demanding applications of models, such as in areas of complex terrain (e.g., hills, valleys, mountains, and over water), when deposition is important, and when atmospheric transformation occurs. As discussed below, it is difficult enough to use models in the simple situations for which they were specifically designed. One should always try to ascertain the level of accuracy that can be expected from a given model in a given application. Sufficient studies have been performed on most air-quality models to address that question.

Zannetti (1990) reviews evaluations of many air-quality models, including Gaussian- plume models. Evaluation procedures have recently been reviewed for photochemical air-quality models (NRC, 1991a). Similar procedures are applicable to other models. In essence, the models should be pushed to their limits, to define the range in which potential errors in either the models themselves or their inputs still lead to acceptable model performances and so that compensatory errors in the models and their inputs (e.g., meteorology, emissions, population distributions, routes of exposure etc.) will be identified. That should lead to a quantitative assessment of model uncertainties and key weaknesses. As pointed out in the NRC (1991a) report, model evaluation includes evaluation of input data. The greatest limitation in many cases is in the availability and integrity of the input data; for the most part, many models can give acceptable results when good-quality input data are available.

A key motivation in model evaluation is to achieve a high degree of confidence in the eventual risk assessment. Pollutant-transport model evaluation, as it pertains to estimating air-pollutant emissions, has been somewhat neglected and is used without adequate discussion and analysis. For example, the modeling of emissions from the ASARCO smelter (EPA, 1985b) showed significant bias. However, the reasons for both the bias and errors were not fully identified. A major plume-model validation study was mounted in the early 1980s with support of the Electric Power Research Institute (EPRI); it was the first study of a large coal-fired power plant situated in relatively simple terrain. The study compared three Gaussian-plume models and three first-order closure numerical (stochastic) models, and an experimental, second-order closure model, for which ground-level concentrations were obtained with both routine and intensive measurement programs (Bowne and Londergan, 1983). (*First-order closure* and *second-order closure* refer to how the effects of turbulence are treated.) The authors conclude that

- The models were poor in predicting the magnitude or location of concentration patterns for a given event.
- The models performed unevenly in estimating peak concentrations as a function of averaging time; none provided good agreement for 1-, 3-, and 24-hour averaging periods.
- The cumulative distribution of hourly concentrations predicted by the models did not match the observed distribution over the full range of concentration values.
- The variation of peak concentration values with atmospheric stability and distance predicted by the Gaussian models did not match the pattern of observed peak values.
- One of the first-order closure models performed better than the Gaussian models in estimating peak concentration as a function of meteorological characteristics, but its predictive

capacity was poorer than desirable for detailed risk assessments, and it systematically over-predicted the distance to the maximal concentrations.

- One of the other first-order closure models systematically underpredicted plume impacts, but its predictive capacity was otherwise superior to that of the Gaussian models.
- An experimental second-order closure model did not provide better estimates of ground-level concentrations than the operational models.

Predictions and observed pollutant concentrations often differed by factors of 2-10. It is clear from the study—in which there was no effect of complex terrain, heat islands, or other complicating effects—that the dispersion models had serious deficiencies. Dispersion models have been developed since then, but they require further development and improvement and they warrant evaluation when applied to new locations or periods.

Larger-scale urban air-quality models perform better in predicting concentrations of secondary species—such as ozone, nitrogen dioxide, and formaldehyde—even though the complex chemical reactions might seem to make the task harder. Prediction accuracy, on the average, is usually within about 10% (NRC, 1990a). This performance is due in part to the coarser spatial resolution used by the model, the chemical transformation times allowing the dispersion from the original sources, and better spatial separation of the sources. The lower spatial resolution, with increased chemical detail and performance, leads back to a consideration of model choice and evaluation: What type of detail is required from a particular model application and what level of performance can be expected?

In summary, model evaluation is an integral part of any risk assessment and is crucial for providing confidence in models. Evaluation procedures have been developed for various classes of air-quality models. Studies have shown that air-quality models can give reasonable predictions, but do not always (or often) do so. Results of a model evaluation can be used in an uncertainty analysis of predicted risk.

Evaluation of EPA Practice

The validity of the population-exposure models used by EPA remains largely untested. Ott et al. (1988) used data from EPA's TEAM studies of carbon monoxide (CO) of Denver and Washington, D.C., to examine the validity of the SHAPE model and compared the estimated co-exposure distribution based on the SHAPE model with the distribution based on direct measurement (personal monitoring). They found the estimated average exposure to be similar with the two approaches, but the ranges in estimated exposure distributions were quite different. The SHAPE exposure model predicted median values well, but there were substantial discrepancies in the tails of the distribution.

Duan (1991) also using data from EPA's TEAM study of carbon monoxide in Washington, D.C., found that the concentrations and time intervals were independent and tested the effectiveness of a "variance-components exposure" model in comparison with SHAPE. Both the long-term average concentrations and short-term fluctuations in concentration were

important in predicting exposure. Duan (1988) and Thomas (1988) examined several statistical parameters for several microenvironments and found the time-invariant component (i.e., a component that does not vary with time, often taken as a background level) to be dominant. Thus, there has been some effort to validate the exposure models developed for research purposes.

There have been no systematic attempts, however, to validate either of the exposure models used for regulatory purposes, the Human Exposure Model (HEM) and the National Ambient Air Quality Standard Exposure Model (NEM). The dispersion-model portion of HEM was compared with other simple Gaussian-plume models, and the results were similar. However, neither actual airborne concentrations nor measured integrated exposures to any airborne constituent were compared with the model results to test its utility in estimating individual or population exposures. Comparison of the site-specific model used to evaluate the health impact of arsenic from the ASARCO smelter in Tacoma, Washington, from the few available data proved to have low marginal accuracy, and arsenic in the exposed human urine samples did not correlate well with estimated exposures, as discussed in Chapter 3. Thus, the effectiveness of these models is essentially unknown, although it will be important to understand their strengths and limitations, including prediction accuracy and the associated uncertainty, when residual risk must be estimated after installation of Maximum Achievable Control Technology (MACT).

When EPA conducts a risk assessment of a hazardous air pollutant, it generally relies on Gaussian-plume models. Gaussian-plume models are inadequately formulated, so inaccuracies appear in predicted pollutant concentrations (e.g., Gaussian-plume models generally are not applicable for nonlinear chemistry or particle dynamics). Furthermore, the inputs to these models are often inaccurate and not directly appropriate for a given application. In practice, application of Gaussian-plume models has not been adequately evaluated, and some evaluations have shown substantial discrepancies. More comprehensive and robust pollutant-transport models (i.e., those more directly applicable to a wider variety of situations) are available, including stochastic Lagrangian and photochemical models, and evaluations have shown good agreement with direct observations. In specific applications, model evaluation (via pollutant monitoring and assessment of model inputs and theory) should be undertaken and ranges of applicability determined. Demonstrations should include, but not be restricted to, showing that the model assumptions reasonably represent physical-chemical behavior of the contaminant, source configuration, and atmospheric dispersion. For environmental conditions for which the performance of Gaussian-plume models are demonstrated to be unsatisfactory, more comprehensive models should be considered; however, their superior performance should be documented and clearly evident when they are considered as alternative in a risk assessment.

EPA has generally not included population activity, mobility, and demographics and has not adequately evaluated the use of population averages (as used by default in HEM) in its exposure assessments. Exposure models, such as NEM and SHAPE, have been developed to account for personal activity. Population-activity models should be used in exposure assessments; however, their accuracy should be clearly demonstrated before considering them as alternatives to the default approach. Demographics might also play a role in determining risk.

Further evaluation of some simple methods (e.g., use of population centroids), compared with more comprehensive tools (e.g., NEM and SHAPE), is warranted, before they are considered in lieu of the default option.

EPA currently uses HEM to screen exposure associated with HAP releases from stationary sources. The HEM-II model uses a standardized EPA Gaussian-plume dispersion model and assumes nonmobile populations residing outdoors at specific locations. The HEM construct is not designed to provide accurate estimates of exposure in specific locations and for specific sources and contaminants when conditions are not represented by the simplified exposure- and dispersion-model assumptions inherent in the standard HEM components. Alternative models for transport and for personal activity and mobility can be adopted in an exposure-modeling system to provide more accurate, scientifically founded, and robust estimates of pollutant-exposure distributions (including variability, uncertainty, and demographic information). Those models can be linked to geographic data bases to provide both geographic and demographic information for exposure-modeling systems.

Application of HEM generally does not include noninhalation exposures to hazardous air pollutants (HAPs) (e.g., dermal exposure), but these routes can be important. Modeling systems similar to extensions of HEM have been developed to account for the other pathways. Unless there is good evidence to the contrary, the contribution of alternative pathways of exposure to HAPs should be considered explicitly and quantified in a risk assessment.

Relatively simple models for exposure assessments, such as HEM, can provide valuable information for setting priorities and determining what additional data should be developed. However, exposure estimates that use this model can have large uncertainties (e.g., a factor of 2-10 due to the Gaussian-plume dispersion model used in HEM alone). Furthermore, Gaussian-plume models, in general, have not been validated for pollutants that are reactive and easily transformed to other chemicals such as organic gases (e.g., formaldehyde), particles, and acids (e.g., nitric and sulfuric acids). Multiple exposure routes can add still more uncertainty as to actual exposure. Uncertainty can be used as a tool for assessing the performance of a model like HEM. This is because HEM is based on very simplified descriptions of pollutant dynamics and was designed for use as a screening tool for estimating human exposure via inhalation.

The predictive accuracy and uncertainty associated with the use of the HEM should be clearly stated with each exposure assessment. The underlying assumption that the calculated exposure estimate is a conservative one should be reaffirmed; if not, alternative models whose performance has been demonstrated to be superior should be used in exposure assessment.

ASSESSMENT OF TOXICITY

The first step in assessing human toxicity based on animal experiments is the extrapolation of observations from studies in rats, mice, monkeys, and other laboratory animals to humans. The extrapolation procedure used in risk assessment to assess the toxicity of a substance is both an intellectual exercise and a tool for making practical decisions. It is based on two assump-

tions: that the biological response to an external stimulus in one species will occur in a different species that is subjected to the same stimulus and that the biological response is proportional to the size of the stimulus (except that a very small stimulus will often result in only a transient response or no immediate response at all). Those two assumptions are invoked whenever extrapolation from animals to humans and from high doses to low doses is performed. Cancer and other end points are discussed separately here because considerations related to extrapolation can differ.

CANCER

QUALITATIVE CONSIDERATIONS

Cancer, defined as abnormal and uncontrolled growth, is ubiquitous among higher organisms; it occurs in plants, animals, and humans. In some cases, carcinogens can be identified as physical or chemical agents or self-replicating infectious agents. Many epidemiological studies have documented an association between exposure to particular chemicals and an increased incidence of particular malignancies in humans (Doll and Peto, 1981). Examples are cancers related to exposure to industrial agents—such as aniline dyes, mustard gas, some metal compounds, and vinyl chloride—and, in the general population, tobacco and tobacco smoke. Perhaps most convincing in this context is the repeated observation that cessation of exposure to a given chemical (e.g., cessation of smoking or introduction of appropriate mitigation or hygienic measures) results in a decrease in cancer incidence. When tested in animal studies, almost all known human carcinogens have been found to produce cancer in other mammals. There are a few exceptions to that rule, e.g. tobacco smoke in laboratory animals. Recent advances in the understanding of basic mechanisms of carcinogenesis, often very similar in laboratory animals and humans, lend credibility to a relationship between animal carcinogenesis and human carcinogenesis, particularly when mutagenicity is involved (OSTP, 1985; Barbacid, 1986; Bishop, 1987); in other cases, advances in the understanding of species-specific mechanisms of carcinogenesis do not support a relationship between humans and specific laboratory animals studied to date (Ellwein and Cohen, 1992). Current long-term carcinogenicity bioassays are conducted with rodents using, among other doses, the highest dose that does not reduce survival as a result of causes other than cancer, known as the maximum tolerated dose (MTD). Information acquired from rodent bioassays conducted at the MTD might yield information on whether a chemical can produce tumors in humans, but it generally cannot provide information on whether it produces tumors through generalized, indirect mechanisms or directly as a result of its specific properties. Mechanistic data could resolve the question of whether it is valid to extrapolate the results of a bioassay to humans (see NRC, 1993b). Current regulatory practice takes the view that in the absence of information to the contrary, animal carcinogens are human carcinogens; however, the data base supporting this assumption is not complete.

Obtaining more information on the biological mechanisms of carcinogenesis, their dose dependence, and their interspecies relevance will permit better and more valid qualitative and quantitative extrapolations. For example, there is a tendency to give more weight to an observation when it relates chemical exposure to development of malignant tumors and to place less emphasis on an observation that suggests that a given chemical induces benign tumors. It might be an oversimplification to consider one category of abnormal growth as invariably detrimental and another as comparatively harmless. Tumor biology is much more complicated. Most, if not all, bronchial adenocarcinomas will kill when they run their course, whereas subcutaneous lipomas will not; however, excision of a malignant basal cell skin tumor is considered a cure, whereas a benign tumor of the VIIIth cranial nerve or of the pituitary gland can be lethal. Available knowledge on causes of cancer and on the biological behavior of tumors does not permit us to ascertain whether a compound that produces a benign tumor in laboratory animals would be either capable or incapable of producing a malignant tumor in humans. In the absence of information to the contrary, the conservative view equates abnormal growth with carcinogenicity. Circumstances that produce benign tumors in animal systems might have the potential for producing abnormal growth in humans, depending on the mechanism involved. Many benign tumors are most easily produced in animal strains that already have an inherently high spontaneous incidence of such tumors (e.g., liver and lung adenomas in mice and mammary tumors in rats). Studies of the genetic, biochemical, hormonal, and other factors that determine development of such tumors might improve the validity of human risk assessments based on animal studies, and should be pursued more vigorously.

The assumption that the organ or tissue affected by a chemical in animals is also the site of greatest risk in humans should also be made cautiously. It is likely that the site of tumor formation is related to the route of exposure and to numerous pharmacokinetic and pharmacodynamic factors. Each route of exposure might result in carcinogenicity and should be considered separately. It probably is reasonable to assume that in some cases, animal models of carcinogenesis can be used to predict the development of human tumors at specific sites, provided that conditions of exposure are comparable. However, if exposure conditions are not similar, that might not be true. For example, it might well be incorrect to assume that agents that produce sarcomas in laboratory animals after subcutaneous injection will induce sarcomas in humans after inhalation. Animal models can be used to detect potential carcinogenicity; however, extrapolating from animal models to particular human organs is not valid without a great deal of additional mechanistic information, such as information on the effects of exposure route, dose, and many other factors, including the metabolism of the agent in question.

EVALUATION OF EPA PRACTICE

Experience has shown that, in a broad sense, extrapolation from species to species is justifiable (Allen et al., 1988; Crump, 1989; Dedrick and Morrison, 1992). It is prudent to

assume that agents that cause abnormal growth of tissue components in laboratory animals will do so in humans. The animal species (mice and rats) most commonly used in the National Toxicology Program (NTP) to make predictions about human carcinogenesis were selected for convenience, not because they have been demonstrated to predict human risks accurately. For example, the risk of inhaled particles for humans might be underestimated in animal assays that use rats and mice, which are obligatory nose breathers and thus might filter out much of the coarser dust. Conversely, some believe that rodents might overpredict human risk when mechanisms of carcinogenesis that are operative in rodents do not occur in humans (Cohen et al., 1992). It appears that NTP has not seriously explored alternatives to rats and mice in carcinogenesis testing, except perhaps for the use of hamsters in inhalation studies.

In principle, selection of data for estimation of carcinogenic potential from the most sensitive strain or species of animals tested is designed to be conservative; whether it is actually conservative and accurate is unknown. This default assumption contributes to the uncertainty in risk assessment, and research designed to investigate the biological mechanisms of carcinogenesis in both rodents and humans should be vigorously pursued so that more accurate risk assessments can be conducted.

QUANTITATIVE CONSIDERATIONS

Key terms in quantitative cancer risk characterization are *unit cancer risk* and *potency*. As currently estimated by EPA, potency is a statistical upper bound on the slope of the linear portion of a dose-response curve at low doses as calculated with a mathematical dose-response model. The unit cancer risk is based on potency and is an upper-bound estimate of the probability of cancer development due to continuous lifetime exposure to one unit of carcinogen. For airborne agents, that unit is commonly defined as exposure to 1 μg of agent per cubic meter of air over a 70-year lifetime.

Cancer potencies are generally based on dose-response relationships generated from cancer bioassays performed with rodents exposed to doses that are several orders of magnitude greater than those for which risk must be estimated. Bioassays typically include two, and to a lesser extent three or more, doses in addition to controls, and are rarely repeated. Often, positive results are obtained at only one dose. Therefore, for most carcinogens, few unequivocal data points are available for potency calculation. In addition, several assumptions often enter into calculations of potency, such as considerations related to tissue dosimetry, in which metabolism data obtained from different experimental systems and used in PBPK modeling might be used in place of bioassay exposure levels. It is not unusual for potency estimates based on the same bioassay data to vary substantially from one risk assessment to another, depending on these additional assumptions and the dose-response model used. Accordingly, potency values are often fraught with as much uncertainty as other aspects of quantitative risk assessment.

To estimate cancer potencies, EPA currently uses the linearized multistage model (EPA, 1987a). This model uses what is essentially an empirical curve-fitting procedure to describe the relationship between bioassay dose and response and to extrapolate the relationship to

exposures below the experimental range. A statistical upper bound on the slope of the low-dose linear portion of the curve is considered to represent an upper bound on a chemical's carcinogenic potency. The multistage model is based on a theory of carcinogenic mechanism proposed in the early 1950s by Armitage and Doll. In essence, normal cells in a target organ are envisioned as undergoing a sequence of irreversible genetic transformations culminating in malignancy. Each transformation to a new stage is assumed to occur at some nonzero background rate. Exposure to a carcinogen is presumed simply to increase one or more of the transformation rates in proportion to the magnitude of the exposure (technically, dose at the target site). However, actual exposure circumstances are more complicated than can be briefly described here. No other potential effects of exposure or alternative mechanisms of carcinogenesis, such as induced cell proliferation or receptor-mediated alterations in gene expression, are included in the Armitage-Doll model. One important consequence of this assumption about how exposure influences transformations is the linearity of risk at low doses, i.e., risk increases and decreases in direct proportion to the delivered dose. That result arises in part because the model assumes that the number of cells at risk of undergoing the first transformation (the susceptible target-cell population) is constant and independent of age, magnitude of exposure, and exposure duration. Thus, the normal processes of cell division, differentiation, and death are not taken into account by the model.

Another cancer dose-response model that has been developed to estimate cancer potencies for risk assessment, but that is not used routinely for regulatory purposes, is the two-stage model. The two-stage model was developed by Moolgavkar, Venzon, and Knudson (Moolgavkar and Venzon, 1979; Moolgavkar and Knudson, 1981; Moolgavkar, 1988; Moolgavkar et al., 1988; Moolgavkar and Luebeck, 1990) and postulates that two critical mutations are required to produce a cancer cell. The model presupposes three cell compartments: normal stem cells, intermediate cells that have been altered by one genetic event, and malignant cells that have been altered by two genetic events. The size of each compartment is affected by cell birth, death, and differentiation processes and by the rates of transition between cell compartments. The model can accomodate some current concepts regarding the roles of inactivated tumor-suppressor genes and activated oncogenes in carcinogenesis. Unlike the Armitage-Doll model, it can explicitly account for many processes considered important in carcinogenesis, including cell division, mutation, differentiation, and death and the clonal expansion of populations of cells. Some knowledge of a chemical's mechanism of action and dose-response data for that mechanism are required to apply the two-stage model, however, and such data on most chemicals are scanty.

Potency estimates are generally based on the assumption that exposure to a particular agent occurs over a 70-year lifetime under constant conditions. That assumption is not likely to apply to the entire exposed population, however, and might produce a conservative estimate of risk. Use of a single potency number implies that the biological response of concern, such as carcinogenesis, depends only on total dose and therefore is independent of dose rate (the quantity of the agent received per unit time). This assumption might be invalid in some cases; for example, studies of low-energy-transfer radiation carcinogenesis show that low-dose-rate

exposures are less effective than high-dose-rate exposures (NRC, 1990b). Other studies of radiation have differing results.

Potency estimates provide a means for comparing animal data with human data and for ranking potential carcinogens. Analysis of data available for some 20 known human carcinogens has shown that, in general, potency values derived from carcinogenicity bioassays in animals agree reasonably well with values calculated for humans from epidemiological studies (Allen et al., 1988). However, ranking of chemicals according to potency should not necessarily be used to make conclusions on the ranking of the corresponding hazards or risk. It is only multiplication of potency (unit risk) with exposure (dose) that yields an estimate of risk. Where there is no exposure, there might be little practical need for information on potency.

Evaluation of EPA Practice

The selection by EPA of a mathematical model to estimate potency is a critical step in quantitative risk assessment, in which alternative assumptions can lead to large differences in estimated risks. Such a model provides explicit, objective rules for extrapolating from the risks observed in controlled, high-dose laboratory experiments to those associated with the far lower doses that people might receive through inhalation. However, all dose-response models are simplified characterizations of the underlying biological reality. That is due, in part, to the incomplete scientific understanding of toxic mechanisms and to the requirement that the models be usable in a broad array of cases.

The challenge for EPA is to incorporate the expanding knowledge of mechanisms into the design of extrapolation models. The models would then depict more accurately the dose-response relationship at the low doses that are of concern to regulators, but are too low for toxic effects to be directly observed in whole animal studies or, often, any feasible human studies. The challenge can be illustrated by examining the simplified mechanistic assumptions that are included in the multistage model used by EPA in light of new understanding of mechanisms, which is not included in that model.

As long as exposure to a chemical has no substantial effect on cell processes other than genetic change, one would not expect the exclusion of these processes from the multistage model to compromise the resulting cancer risk estimates. The model would likely be appropriate for "direct-acting" carcinogens—ones, such as radiation, that act by directly attacking cellular DNA and thereby causing genetic transformation. In recent years, however, it has become apparent that many substances alter the pharmacodynamics of cells and can be carcinogenic by mechanisms that do not involve direct covalent interaction with DNA at all, but involve indirectly caused alterations in gene expression. One consequence of such a change could be altered cellular dynamics in the target organ. Because genetic transformations can occur spontaneously, many target organs contain a background of continuing steps in the multistep carcinogenic process. Exposure to a chemical could augment those background carcinogenic processes by simply increasing the pool of cells that are susceptible to further

transformation. Such augmentation might occur as a regenerative response to cellular injury among surviving cells or to the cell-killing that occurs after exposure to highly toxic substances. The augmentation of background carcinogenic processes could also occur as an indirect response to alterations in hormonal balances induced by exposure or as a response to a directly mitogenic substance, i.e., one that stimulates normal cell division. By increasing the rate of cell division, such substances can increase the overall probability of generating a mutation, even though they have no direct effect on the transformation probability per cell division.

Similarly, exposure to substances classified as nongenotoxic carcinogens or "promoters" can create physiologic conditions within a target organ that favor the growth of "initiated" cells, i.e., cells that have already sustained at least one irreversible change from normal cells. Clonal expansions of initiated cell populations can be induced by exposure to promoters, thus increasing the probability of cell transformation and malignancy without directly affecting DNA.

Critical to effective regulatory use of biologically based models such as the two-stage model is accurate determination of the dose-response and time-response relationships for agent-induced cell death, differentiation, transformation and division, if any, in target tissues. Those processes might exhibit threshold-like dose-response relationships, in contrast to the presumed low-dose linear response of conventional multistage model transformation rates. Conversely, better understanding might show supralinear relations. Thus, use of a two-stage pharmacodynamic model might predict low-dose risks that are lower or higher than those predicted by the linearized multistage model.

Successful use of biologically based models in the risk assessment process will require a greater variety and amount of information on and understanding of carcinogenic mechanisms than is typically available for most chemicals. In the near term, such a data-intensive approach might be applied only to substances that have great economic value. In the long run, as knowledge and experience accrue, the use of models that incorporate relevant pharmacodynamic data should become more routine. Those models, used in conjunction with pharmacokinetic models for determining delivered doses, will increase the accuracy of quantitative risk assessment. For that reason, EPA should intensify their incorporation into the cancer-risk assessment process. For more information on two-stage models, see the NRC (1993c) report on this topic.

CARCINOGEN CLASSIFICATION

As noted in Chapter 4 (Table 4-1), EPA, following the lead of the International Agency for Research on Cancer (IARC), provides an evaluation of the available evidence of carcinogenicity of individual substances. The direction and strength of evidence are summarized by a letter: A, B_1, B_2, C, D, or E (see Table 4-1). The assignment of a substance to a class (actually, the assignment of available evidence to a class) depends almost entirely on epidemiological evidence and evidence derived from animal studies. The evidence for each of these is

classified by EPA as "sufficient," "inadequate," or "limited." Some other types of experimental evidence (e.g., on genotoxicity) might sometimes play a role in the classification, but the epidemiological and bioassay data are generally of overriding importance.

The EPA classification scheme is intended to provide information on hazard—not to provide information about potential human risk; the latter cannot be assessed without the additional evaluation of dose-response and exposure information. The assignment of evidence to a class is intended by EPA only to suggest how convinced we should be that a substance poses a carcinogenic hazard to people. The classification is thus meant to depict the state of our knowledge regarding human carcinogenic hazard.

The difference between hazard and risk needs to be further emphasized here. As conceived in EPA's current four-step approach, identifying a substance as a possible, probable, or known carcinogenic hazard to humans means only that, under some unspecified conditions, the substance could cause excess cancers to occur in people. Evaluation of potency and of the exposures incurred by specific populations provides the information needed to assess the probability (risk) that the substance will cause cancer in the specified population. EPA developed the categorization scheme because it believes that, in addition to the risk estimate, decision-makers should have some sense of the strength of the evidence supporting identification of a substance as a carcinogen. There has been some confusion regarding the terms *strength of evidence*, as used by EPA, and *weight of evidence*. Some interpret *strength* to only describe the degree of positive evidence and *weight* to apply when *all* evidence—positive, negative, and evidence on relevance to humans—is considered. The committee adopts those uses of the terms. In many cases, substances for which the evidence of human carcinogenicity is strong (classification A) will, in specific circumstances, pose relatively small risks (because of low potency or low exposure), whereas substances for which the evidence of human carcinogenicity is much less convincing (classification B_2, for example) are likely to pose large risks (because of high potency or exposure). The typical question faced by a decision-maker is whether, for example, more restrictive controls should be placed on substances in class A that pose relatively small risks or on substances in lower classes that pose equal or greater risks. Stated in other terms, the issue concerns the justification for placing different degrees of regulatory restriction on substances that pose equal risks but which are differently classified. Should we control more carefully substances for which the state of our knowledge regarding human carcinogenicity is highly certain than we do substances for which the state of our knowledge is relatively weak? Although EPA includes a strength-of-evidence classification with each risk characterization, there is no clear indication of whether and how the classification influences ultimate agency decision-making.

EVALUATION OF EPA PRACTICE

Does EPA's approach accurately portray the state of knowledge regarding human carcinogenic hazard? It is certainly the case that the state of scientific knowledge regarding the potential for various substances to contribute to the development of human cancers is highly

variable among them. It also seems reasonable that risk assessors should have available a means to express that knowledge in a relatively simple way. It is for this reason that any such scheme should be examined carefully to ensure that it expresses as closely as possible what it is intended to express and that it summarizes all the relevant and appropriate findings derived from data, with no extraneous data.

Because two conclusions (that the substance might pose a carcinogenic hazard to humans under some conditions of exposure and that animal data can be unconditionally extrapolated to humans) are implicitly contained in the current EPA classification system, it could be conceived as misleading in some cases in which the scientific evidence does not support one or more of the typical default assumptions (for example, on route-to-route, high to low dose, or animal-to-human extrapolation). Such a situation could arise when, for example, data are available to show clearly and convincingly that some types of animal tumors would not likely to be produced in humans or when mechanistic data show that results obtained at high doses are not relevant to low doses. Although different in kind, classification of substances at EPA's D or E level could also be misleading. If, for example, a substance were classified at level E on the basis of negative chemical bioassays in two species, but additional data suggested that neither animal species metabolized the substance in the way humans did, then the absence of potential human hazard would be improperly inferred.

The present EPA system might also be misleading because it is too susceptible to "accidents of fate." The carcinogenicity of a substance that happens to cause very rare tumors in humans (e.g., vinyl chloride, which causes angiosarcoma of the liver) is much easier to detect in epidemiological studies than is the carcinogenicity of a substance that causes very common human cancers, such as colorectal carcinoma. Although the available animal data on the latter substance might be very convincing with respect to carcinogenicity and there might be every reason to believe that it will be as hazardous to humans as the former (i.e., the "known" human, category A carcinogen), it will usually end up in category B, which may be interpreted as suggesting a lesser likelihood of hazard. Such a distinction might be due only to differences in our ability to detect the carcinogenic properties of substances that produce different types of cancers, and not to any true differences in human hazard.

POSSIBLE IMPROVEMENTS IN EPA PRACTICE

Before turning to the issue of improvements in EPA's carcinogen classification scheme, the committee first considered whether any such scheme should be used at all. As noted above, the current scheme can easily be misinterpreted—unfamiliar users might be led to believe that all substances in a specific category are equally hazardous or nonhazardous. Moreover, it is impossible to capture in any simple categorization scheme the completeness and complexity of the information that supports scientific judgments about the nature of a human carcinogenic hazard and the conditions under which it can exist. The quality, nature, and extent of such information vary greatly among carcinogens, and it is not an exaggeration to state that every substance is unique with respect to the scientific evidence bearing on its hazards.

It is for these reasons that the committee strongly recommends that EPA include in each hazard-identification portion of a risk assessment a *narrative* evaluation of the evidence of carcinogenicity. Such a narrative should contain at least the following:

- An evaluation of the strength of the available human and animal evidence.
- A weight-of-evidence evaluation of any available information on the relevance to humans of the animal models used and the results obtained from them and on the conditions of exposure (route, dose, duration, and timing) under which carcinogenic responses to other conditions of exposure (usually conditions that could exist in human populations exposed environmentally) have been measured (either in human populations or in laboratory animals).

Such a narrative seems to be the best way to describe the type of information typically available to evaluate carcinogenic hazards and should be used by EPA when it undertakes full-scale risk assessments.

Although the committee agreed that such narrative descriptions are the preferred way to express scientific evidence, it also recognized that there are important practical needs for some type of simple categorization of evidence. The committee recognized, for example, that many regulatory actions or plans for action require, for practical reasons, the creation of lists of carcinogens and that narrative statements are not likely to be included in such lists. Without some simple categorization scheme, such lists are likely to be completely undiscriminating with respect to the potential human hazards of the substances on them. When any such lists are used, for example, to create priorities for full risk assessment or for some type of regulation, the results could be seriously misleading to decision-makers and the public.

As already noted, however, the committee believes that the current EPA categorization scheme is inadequate. Substantial improvements could be made if the scheme incorporated not only "strength-of-evidence" information, but also some of the information we have called for in the narrative description.

It will not be easy to create a categorization scheme for carcinogens that incorporates both strength of evidence and the two "relevance" considerations. Moreover, EPA is not the only agency for which such a categorization scheme is useful. Indeed, there is a strong need for international agreement on a single classification. It would be highly desirable for EPA to convene a workshop on the matter and involve other agencies of federal and state governments, IARC, and other national and international bodies to develop a scheme that would have worldwide acceptance. IARC has recently moved to include information on mechanisms of carcinogenic action in its evaluation of carcinogens. Such an effort seems essential to eliminating the deficiencies of current schemes and the confusion that exists because of differences in approaches to categorization around the globe.

The committee suggests the scheme in Table 7-1 as a draft or prototype to avoid the difficulties of the current EPA scheme. The proposal in this table incorporates both strength-of- evidence considerations (as in the current EPA and IARC schemes) and "relevance" information, as specified in the two points mentioned above. The example also reduces the

TABLE 7-1. continued

Step 2: Categorization according to strength of evidence (a through d, in decreasing order of strength)

Category	Data Source	Subcategory			
		a	b	c	d
I	Epidemiology Animal Studies	S S/L/NI	L S	NI S	NI L
II	Epidemiology Animal Studies	s s/l/NI	l s	NI s	NI l
III	Epidemiology Animal Studies	L/NI s_l	NI l_l		
IV	Epidemiology Animal Studies	NI NI	NI/NA NA		

S = sufficient evidence, high relevance. s = sufficient evidence, limited relevance.
L = limited evidence, high relevance. l = limited evidence, limited relevance.
NI = no or inadequate evidence. s_l = sufficient evidence, low relevance.
NA = no evidence in adequate studies. l_l = limited evidence, no relevance.

susceptibility of current classification schemes to the "accidents of fate" that can artificially influence the availability of evidence for different substances.

The classification in Table 7-1 takes place in two steps. In Step 1, a classification is made (into Categories I-IV) according to the two relevance criteria mentioned above. Note also that Category I is used for all substances on which positive carcinogenicity data are available and on which there are no substantive data to support conclusions that would place them in Category II or III—i.e., *Category I is the default option that applies when data related to relevance are weak or absent.* Step 2 of the classification involves evaluation of the strength of the available evidence.

Such a categorization scheme can provide guidance on priorities for both risk assessment and a variety of regulatory efforts. Substances placed in Category I, for example, would generally receive greater attention with respect to their carcinogenic properties than those in Category II; and within Category I, the nature of the attention received might be further influenced by the strength of available evidence (i.e., Ia > b > c > d). A Ia substance, for example, might be a prime candidate for immediate and stringent regulation, whereas a Id substance might be a prime candidate for high-priority information-gathering.

Placement of a substance in Category II does not mean that regulatory efforts should not be

TABLE 7-1. Possible scheme for categorizing carcinogens.

Step 1: Categorization according to relevance of findings to humans

Category	Nature of Evidence
Category I Might pose a carcinogenic hazard to humans under any conditions of exposure. Magnitude of risk depends on dose-response relationship and extent of human exposure.	• Evidence of carcinogenicity in either human or animal studies (strength of evidence varies; see Step 2) • No information available to raise doubts about the relevance to humans of animal model or results • No information available to raise doubts about relevance of conditions of exposure (route, dose, timing, duration, etc.) under which carcinogenic effects were observed to conditions of exposure likely to be experienced by human populations exposed environmentally.
Category II Might pose a carcinogenic hazard to humans, but only under limited conditions. Whether a risk exists in specific circumstances depends on whether those conditions exist. Dose-response and exposure assessments must be completed to identify conditions under which risk exists.	• Evidence of carcinogenicity in either human or animal studies (strength of evidence varies; see Step 2) • Scientific information available to show that there are limitations in the conditions under which carcinogenicity might be expressed, owing to questions about the relevance to humans of the animal models or results or relevance of the conditions of exposure (route, dose, timing, duration, etc.) under which carcinogenic effects were observed to conditions of exposure likely to be experienced by human populations exposed environmentally.
Category III Notwithstanding the evidence of carcinogenicity in animals, not likely to pose a carcinogenic hazard to humans under any conditions.	• Evidence of carcinogenicity in animal studies • Scientific information available to show that the animal models or results are not relevant to humans under any conditions.
Category IV Evidence available to demonstrate lack of carcinogenicity or no evidence available.	• No evidence of carcinogenicity or evidence of non-carcinogenicity (weight of negative evidence varies; see Step 2)

undertaken. For example, there might be reason to determine whether potentially risky conditions of exposure exist in any situations. The categories do not influence ultimate actions, but only priorities and the relative, inherent degrees of concern associated with different substances.

Although the committee recommends that any categorization scheme adopted by EPA include the elements associated with the above example, it also recognizes that there might be other ways to capture and express the same information. Some members suggested, for example, that substances listed as carcinogens simply be accompanied by a set of codes that specify both the strength of supporting evidence and the conditions and limitations, if any, that might pertain to the interpretation of that evidence (e.g., an asterisk next to a chemical might mean "assumed to be carcinogenic in humans only when inhaled").

Other End Points of Toxicity

The standard approach to regulating chemicals that are associated with noncancer end points of toxicity has been based on the theory of homeostasis. According to that theory, biological processes that maintain homeostasis exist in an interdependent web of adaptive responses that automatically react to and compensate for stimuli that alter optimal conditions. An optimal condition is maintained as long as none of the stimuli that regulate it is pushed beyond some limit or "threshold." For the purposes of regulation, end points of toxicity other than cancer are lumped together under a toxicological paradigm that presumes a dose threshold for any chemical capable of inducing an adverse effect: there is an exposure below which the adverse effect would not be expected to occur. The current approach—no-observed-adverse-effect level (NOAEL) and uncertainty factor—is only a semiquantitative method designed to prevent exposures that are likely to result in an adverse effect, not a mechanistically based quantitative method for assessing the likely incidence and severity of effects in an exposed population. Moving beyond the current simplistic regulatory method will require, as is the case for carcinogenesis, a greater understanding of the mechanisms of disease causation, of pharmacokinetics, and of interindividual variation in each. Such improved understanding will permit final abandonment of the obsolete "threshold versus nonthreshold" paradigm for regulating carcinogens and noncarcinogens.

Evaluation of EPA Practice

The methodology now used by EPA to regulate human exposure to noncarcinogens is in a state of flux. That used by EPA in the past was not sufficiently rigorous. It was not based on evaluations of biological mechanisms of action or on differences in susceptibility between and within exposed populations. In addition, it incorporated risk management, not scientifically based risk-assessment techniques; and it did not permit incorporation of newer and better scientific information as it was obtained. The NOAEL-uncertainty factor approach might be

adequate for the immediate future as a screening technique and for setting priorities, but its empirical and scientific basis is meager. EPA appears to be continuing to pursue simplistic, empirical techniques by adding to the list of uncertainty factors in use.

Impact of Pharmacokinetic Information in Risk Assessment

One of the critical steps in risk assessment is the selection of the measure of exposure to be used in defining the dose-response relationship. It is common today to calculate exposure on the basis of the "administered dose" of a chemical—the dose or amount fed to animals in toxicity studies or ingested by humans in food or water or inhaled in air. That dose can usually be accurately measured.

The dose that is of interest for risk assessment, however, is the amount of the biologically active form of a substance that reaches specific target tissues. This target-tissue dose is the "delivered dose," and its biologically active derivative, if any, is the "biologically active dose." The biologically active dose causes the events that culminate in toxicity to target cells and organs, and ideally it is used as the basis for defining the dose-response relationship and for assessing risk. The science of pharmacokinetics seeks to replace the current operating assumption—that administered dose and delivered dose are always directly proportional and that the administered dose is therefore an appropriate basis for risk assessment—with direct, accurate information about the delivered or biologically active dose.

Pharmacokinetic models are used to study the quantitative relationship between administered and delivered or biologically active doses. The relationship reflects the spectrum of biological responses to exposure, from physiological responses of a whole organism to biochemical responses within specific cells of a target organ. Pharmacokinetic models explicitly characterize biologic processes and permit accurate predictions of the doses of an agent's active metabolites that reach target tissues in exposed humans. As a consequence, the use of pharmacokinetic models to provide inputs to dose-response models reduces the uncertainty associated with the dose parameter and can result in more accurate estimates of potential cancer risks in humans.

The relationship between administered and delivered doses often differs among individuals: because of such differences, some people might be acutely sensitive and others insensitive to the same administered dose. The relationship between administered and delivered doses can also differ between large and small exposures and between continuous and intermittent exposures, and it can differ among species, some species being more or less efficient than humans in the transport of an administered dose to tissues or in its metabolism to a biologically active or inactive derivative. Those differences in the relationship between administered and delivered or biologically active doses can dramatically affect the validity of the predictions of dose-response models; failure to incorporate the difference into the models contributes to the uncertainty in risk assessment.

Differences between administered and biologically active doses occur because specialized organ systems intervene to modulate the body's responses to inhaled, ingested, or otherwise

absorbed toxic materials. For example, the liver can detoxify materials circulating in the blood by producing enzymes to accelerate chemical reactions that break the materials down into harmless components (metabolic deactivation, or "detoxification"). Conversely, some substances can be activated by metabolism into more toxic reaction products. Activation and detoxification might occur at the same time and can occur in the same or different organ systems.

Furthermore, the rates at which activation and detoxification take place might have natural limits. Metabolic deactivation might thus be overwhelmed by high exposure concentrations, as seems to be the case with formaldehyde: the biologically active dose and the risk of nasal-tumor development rise rapidly in exposed rats only at high airborne concentrations. The assumption of a simple linear relationship between administered and biologically active doses of formaldehyde is believed by many to result in exaggerated estimates of cancer risk at low exposure concentrations. In contrast, metabolic activation of vinyl chloride occurs more and more slowly with increasing administered dose, because a critical enzyme system becomes overloaded; the biologically active dose and the resulting liver-tumor response increase more and more slowly as the administered dose increases. The assumption of a linear relationship between administered and delivered doses in the case of vinyl chloride could result in underestimation of the cancer risk associated with low doses. These examples illustrate how using pharmacokinetic models can reduce the uncertainty in risk estimation by modifying the dose values used in dose-response modeling to reflect the nonlinearity of metabolism.

Although most pharmacokinetic models are derived from laboratory-animal data, they provide a biological framework that is useful for extrapolating to human pharmacokinetic behavior. Anatomical and physiological differences among species are well documented and easily scaled by altering model parameters for the species in question. This aspect of pharmacokinetic modeling reduces the uncertainty associated with extrapolating from animal experiments to human cancer risk. For example, considerable effort has been devoted to the development of pharmacokinetic models for methylene chloride, which is considered a rodent carcinogen. The model was initially developed on the basis of rat data, then scaled to predict human behavior. Predictions in humans were compared with published data and with the results of experiments in human volunteers. The model was shown to predict accurately the pharmacokinetic behavior of inhaled methylene chloride and its metabolite carbon monoxide in both species (Andersen et al., 1991). Use of a particular pharmacokinetic model for methylene chloride in cancer risk assessment reduces human risk estimates for exposure to methylene chloride in drinking water by a factor of 50-210, compared with estimates derived by conventional linear extrapolation and body surface-area conversions (Andersen et al., 1987). Other analyses show different results (Portier and Kaplan, 1989). What pharmacokinetic models for methylene chloride do not predict, however, is whether methylene chloride is a human carcinogen. Thus, although use of the model might improve confidence in dose estimation by replacing the conventional scaling-factor approach, it cannot predict the outcome of exposure in humans.

Another way to reduce uncertainty would be to use pharmacokinetic models to extrapolate between exposure routes. If information on the disposition of an agent were available only as

a result of its inhalation in the workplace, for example, and a risk assessment were required for its consumption in drinking water, appropriate models could be constructed to relate the delivered dose after inhalation to that expected after ingestion. To the committee's knowledge, pharmacokinetic models have not yet been used in a risk assessment for such regulatory purposes.

Failure to include pharmacokinetic considerations in dose-response modeling contributes to the overall uncertainty in a risk assessment, but uncertainty is associated with their use as well. This uncertainty comes from several sources. First, uncertainty is associated with the pharmacokinetic model parameters themselves. Parameter values are usually estimated from animal data and can come from a variety of experimental sources and conditions. Quantities can be measured indirectly, they can be measured *in vitro*, and they can vary among individuals. Different data sets might be available to estimate values of the same parameters. Hattis et al. (1990) evaluated seven pharmacokinetic models for tetrachloroethylene (perchloroethylene) metabolism and found that their predictions varied considerably, primarily because of the differences in choice of data sets used to estimate values of model parameters. Moreover, analogous parameter values are also needed for humans—although some values, such as organ weights, are amenable to direct measurement and do not vary widely among humans, others, such as rate constants for enzymatic detoxification and activation, are both difficult to measure and highly variable.

Second, there is uncertainty in the selection of the appropriate tissue dose available to model. For example, information might be available on the blood concentration of an agent, on its concentration in a tissue, or on the concentrations of its metabolites in the tissue. Tissue concentrations of one metabolite might be inappropriate if another metabolite is responsible for the biologic effects. Total tissue concentrations might not accurately reflect the biologically active dose if only one type of cell within the tissue is affected.

Choice of an appropriate measure of tissue dose can have an effect on cancer risk estimates. Farrar et al. (1989) considered three measures of tissue dose for tetrachloroethylene: tetrachloroethylene in liver, tetrachloroethylene metabolites in liver, and tetrachloroethylene in arterial blood. Using EPA's pharmacokinetic model for tetrachloroethylene and cancer bioassay data in mice, they found that human cancer risk estimates varied by a factor of about 10,000, depending on the dose surrogate used. Interestingly, the estimates bracketed that obtained in the absence of any pharmacokinetic transformation of dose as shown in Table 7-2.

This example illustrates the variation in dose and risk estimates that can be obtained under different assumptions, but it does not help to evaluate of the validity of any of the estimates in the absence of knowledge of the biologic mechanism of action of tetrachloroethylene as a rodent carcinogen and in the absence of knowledge of whether it is a human carcinogen. Although the dose of metabolites to the liver appears to be the most appropriate choice of dose surrogate, there is a high degree of nonlinearity between this dose and the tumor incidence in mice. The nonlinearity indicates either that this dose surrogate does not represent the actual

TABLE 7-2. Risk estimates based on EPA's pharmocokinetic model for tetrachlorethylene and cancer bioassay data in mice.

Dose Surrogate	Risk Estimate[a]
Administered dose	5.57×10^{-3}
Dose to liver	425×10^{-3}
Dose of metabolites to liver	0.0195×10^{-3}
Dose in blood	126×10^{-3}

[a]Maximum-likelihood estimate.

Source: Farrar et al., 1989.

biologically active dose for the particular sex-species combination analyzed by these authors or that the model does not adequately describe tetrachloroethylene pharmacokinetics.

The science of pharmacokinetics seeks to gain a clear understanding of all the biological processes that affect the disposition of a substance once it enters the body. It includes the study of many active biological processes, such as absorption, distribution, metabolism (whether activation or deactivation), and excretion. Accurate prediction of delivered and biologically active doses requires comprehensive, physiologically based computer models of those linked processes. Because the science of pharmacokinetics aims to replace general assumptions with a more refined model based on the specific relationship between administered and delivered or biologically active doses, its use in risk assessment will help to reduce the uncertainties in the process and the related bias in risk estimation. Advances will come slowly and at considerable cost, because detailed knowledge of the biologically active dose of many materials must be acquired before generalizations can be confidently exploited. Nevertheless, EPA increasingly incorporates pharmacokinetic data into the risk-assessment process, and its use represents one of the clearest opportunities for improving the accuracy of risk assessments.

CONCLUSIONS

Developing improved methods for assessing the long-term health impacts of chemicals will depend on improved understanding of the underlying science and on more effective coordination, validation, and integration of the relevant environmental, clinical, epidemiological, and laboratory data, each of which is limited by various kinds of error and uncertainty. Goodman and Wilson (1991) have demonstrated that, for 18 of 22 chemicals studied, there is good

agreement between risk estimates based on rodent data and on epidemiologic studies. Their quantitative assessment, which can be compared to the Ennever et al. (1987) qualitative evaluation of the same issue, provides stronger evidence that current risk-assessment strategies produce reasonable estimates of human experience for known human carcinogens (Allen et al., 1988).

The reliability of a given health-risk assessment can be determined only by evaluating both the validity of the overall assessment and the validity of its components. Because the validity of a risk assessment depends on how well it predicts health effects in the human population, epidemiologic data are required for testing the predictions. To the extent that the requisite data are not already available, epidemiologic research will be necessary. An example is the study in which the New York Department of Health conducted biological monitoring for arsenic in schoolchildren (New York Department of Health, 1987). The researchers compared their findings with the arsenic concentrations predicted by the risk assessment conducted by EPA. The good agreement between the estimates and actual urinary arsenic concentrations in the children provided support for the EPA risk model.

The committee believes that substantial research is warranted to validate methods, models, and data that are used in risk assessment. In some instances the magnitude of uncertainty is not well understood, because information on the accuracy of the prediction process for each model used in risk assessment is insufficient. We also note that the uncertainties tend to vary considerably; for example, uncertainties are relatively low for estimation of population characteristics, compared with those associated with extrapolation from rodents to human beings.

The quality of risk analysis will improve as the quality of input improves. As we learn more about biology, chemistry, physics, and demography, we can make progressively better assessments of the risks involved. Risk assessment evolves continually, with re-evaluation as new models and data become available. In many cases, new information confirms previous assessments; in others, it necessitates changes, sometimes large. In either case, public confidence in the process demands that EPA make the best judgments possible. That an estimate of risk is subject to change is not a criticism of the process or of the assessors. Rather, it is a natural consequence of increasing knowledge and understanding. Re-evaluating risk assessments and making changes should be expected, embraced, and applauded, rather than criticized.

FINDINGS AND RECOMMENDATIONS

The following is a compilation of findings and recommendations related to evaluation of methods, data, and models for risk assessment.

Predictive Accuracy and Uncertainty of Models

Various methods and models are available to EPA and other organizations for conducting emission characterization, exposure assessment, and toxicity assessments. They include those used as default options and their corresponding alternatives, which represent deviations from the defaults. The predictive accuracy and uncertainty of the methods and models used for risk assessment are not clearly understood or fully disclosed in all cases.

- EPA should establish the predictive accuracy and uncertainty of the methods and models and the quality of data used in risk assessment with the high priority given to those which support the default options. EPA and other organizations should also conduct research on alternative methods and models that might represent deviations from the default options to the extent that they can provide superior performance and thus more accurate risk assessments in a clear and convincing manner.

Emission Characterization

Guidelines

EPA does not have a set of guidelines for emission characterization to be used in risk assessment.

- EPA should develop guidelines that require a given quality and amount of emission information relative to a given risk-assessment need.

Uncertainty

EPA does not adequately evaluate the uncertainty in the emission estimates used in risk assessments.

- Because of the wide variety of processes and differing maintenance of those sources, EPA should develop guidelines for the estimation and reporting of uncertainty in emission estimates; these guidelines may depend on the level of risk assessment.

External Collaboration

EPA has worked with outside parties to design emission characterization studies that have moved the agency from crude to more refined emission characterization.

• EPA should conduct more collaborative efforts with outside parties to improve the overall risk-assessment process, and each step within that process.

EXPOSURE ASSESSMENT

GAUSSIAN-PLUME MODELS

In its regulatory practice, EPA has relied on Gaussian-plume models to estimate the concentrations of hazardous pollutants to which people are exposed. However, Gaussian-plume models are crude representations of airborne transport processes; because they are not always accurate, they lead to either underestimation or overestimation of concentrations. Stochastic Lagrangian and photochemical models exist, and evaluations have shown good agreement with observations. Also, EPA has typically evaluated its Gaussian-plume models for release and dispersion of criteria pollutants from plants with good dispersion characteristics (i.e., high thermal buoyancy, high exit velocity, and tall stacks). EPA has not fully evaluated the Gaussian-plume models for hazardous air pollutants with realistic plant parameters and locations; thus, their potential for underestimation or overestimation has not been fully disclosed.

• EPA should evaluate the existing Gaussian-plume models under more realistic conditions of small distances to the site boundaries, complex terrain, poor plant dispersion characteristics (i.e., low plume buoyancy, low stack exit momentum, and short stacks), and presence of other structures in the plant vicinity. When there is clear and convincing evidence that the use of Gaussian-plume models leads to underestimation or overestimation of concentrations (e.g., according to monitoring data), EPA should consider incorporating state-of-the-art models, such as stochastic-dispersion models, into its set of concentration-estimation models and include a statement of criteria for their selection and for departure from the default option.

EXPOSURE MODELS

EPA has not adequately evaluated HEM-II for estimation of exposures, and prior evaluations of exposure models have shown substantial discrepancies between measured and predicted exposures, i.e., yielding under prediction of exposures.

• EPA should undertake a careful evaluation of all its exposure models to demonstrate their predictive accuracy (via pollutant monitoring and assessment of model input and theory) for estimating the distribution of exposures around plants that limit hazardous air pollutants. EPA should particularly ensure that, although exposure estimates are as accurate as possible, the exposure to the surrounding population is not underestimated.

POPULATION DATA

EPA has not previously used population activity, population mobility, and demographics in modeling exposure to hazardous air pollutants and has not adequately evaluated the effects of assuming that the population of a census enumeration district is all at the location of the district's population center.

- EPA should use population-activity models in exposure assessments when there is reason to believe that the exposure estimate might be inaccurate (e.g., as indicated by monitoring data) if the default option is applied. This is particularly important in the case of potential underestimation of risk. Population mobility and demographics will also play a role in determining risk and lifetime exposures. EPA should conduct further evaluation of the use of both simple methods (e.g., use of center of the population examined) and more comprehensive tools (e.g., NEM and SHAPE exposure models).

HUMAN-EXPOSURE MODEL

EPA uses the Human-Exposure Model (HEM) to evaluate exposure associated with hazardous air-pollutant releases from stationary sources. This model generally uses a standardized EPA Gaussian-plume dispersion model and assumes nonmobile populations residing outdoors at specific locations. The HEM construct will not provide accurate estimates of exposure in specific locations and for specific sources and contaminants where conditions do not match the simplified exposure and dispersion-model assumptions inherent in the standard HEM components.

- EPA should provide a statement on the predictive accuracy and uncertainty associated with the use of the HEM in each exposure assessment. The underlying assumption that the calculated exposure estimate based on the HEM is a conservative one should be reaffirmed; if not, alternative models whose performance has been clearly demonstrated to be superior should be used in exposure assessment. These alternative models should be adapted to include both transport and personal activity and mobility into an exposure-modeling system to provide more accurate, scientifically founded, and robust estimates of pollutant exposure distributions (including variability, uncertainty, and demographic information). Consideration may be given to linking these models to geographic information systems to provide both geographic and demographic information for exposure modeling.

EPA generally does not include non-inhalation exposures to hazardous air pollutants (e.g., dermal exposure and bioaccumulation); its procedure can lead to underestimation of exposure. Alternative routes can be an important source of exposure. Modeling systems similar to extensions of HEM have been developed to account for the other pathways.

• EPA should explicitly consider the inclusion of noninhalation pathways, except where there is prevailing evidence that noninhalation routes—such as deposition, bioaccumulation, and soil and water uptake—are negligible.

ASSESSMENT OF TOXICITY

EXTRAPOLATION FROM ANIMAL DATA FOR CARCINOGENS

EPA uses laboratory-animal tumor induction data, as well as human data, for predicting the carcinogenicity of chemicals in humans. It is prudent and reasonable to use animal models to predict potential carcinogenicity; however, additional information would enhance the quantitative extrapolation from animal models to human risks.

• In the absence of human evidence for or against carcinogenicity, EPA should continue to depend on laboratory-animal data for estimating the carcinogenicity of chemicals. However, laboratory-animal tumor data should not be used as the exclusive evidence to classify chemicals as to their human carcinogenicity if the mechanisms operative in laboratory animals are unlikely to be operative in humans; EPA should develop criteria for determining when this is the case for validating this assumption and for gathering additional data when the finding is made that the species tested are irrelevant to humans.

EPA uses data that generally assume that exposure of rats and mice after weaning and until the age of 24 months is the most sensitive and appropriate test system for conservatively predicting carcinogenicity in humans. These doses miss exposure of animals before they are weaned including newborns. Furthermore, the sacrifice of animals at the age of 2 years makes it difficult to estimate accurately the health affects of a disease whose incidence increases with age (as does that of cancer).

• EPA should continue to use the results of studies in mice and rats to evaluate the possibility of chemical carcinogenicity in humans. EPA and NTP are encouraged to explore the use of alternative species to test the hypothesis that results obtained in mice and rats are relevant to human carcinogenesis, the use of younger animals when unique sensitivity might exist for specific chemicals, and the age-dependent effects of exposure.

EPA typically extrapolates data from laboratory animals to humans by assuming that the delivered dose is proportional to the administered dose, as a default option. Alternative pharmacokinetic models are used less often to link exposure (applied dose) to effective dose.

• EPA should be encouraged to continue to explore and, when it is scientifically appropriate, incorporate mechanism-based pharmacokinetic models that link exposure and biologically effective dose.

The location of tumor formation in humans is related to route of exposure, chemical properties, and pharmacokinetic and pharmacodynamic factors, including systemic distribution of chemicals throughout the body. Thus, tumors might be found at different sites in humans and laboratory animals exposed to the same chemical. EPA has accepted evidence of carcinogenicity in tissues of laboratory animals as evidence of human carcinogenicity without necessarily assuming correspondence on a tumor-type or tissue-of-origin basis. EPA has extrapolated evidence of tumorigenicity by one route to another route where route-specific characteristics of disposition of the chemical are taken into account. EPA has traditionally treated almost all chemicals that induce cancer in a similar manner, using a linearized multistage nonthreshold model to extrapolate from large exposures and associated measured responses in laboratory animals to small exposures and low estimated rates of cancer in humans.

- Pharmacokinetic and pharmacodynamic data and models should be validated, and quantitative extrapolation from animal bioassays to humans should continue to be evaluated and used in risk assessments. EPA should continue to use the linearized multistage model as the default for extrapolating from high to low doses. If information on the mechanism of cancer induction suggests that the slope of the linearized multistage model is not appropriate for extrapolation, this information should be made an explicit part of the risk assessment. If sufficient information is available for an alternative extrapolation, a quantitative estimate should be made. EPA should develop criteria for determining what constitutes sufficient information to support an alternative extrapolation. The evidence for both estimates should be made available to the risk manager.

Extrapolation of Animal Data on Noncarcinogens

EPA uses a semiquantitative NOAEL-uncertainty factor approach to regulating human exposure to noncarcinogens.

- EPA should develop biologically based quantitative methods for assessing the incidence and likelihood of noncancer effects in an exposed population. These methods should permit the incorporation of information on mechanisms of action, as well as on differences in population and individual characteristics that affect susceptibility. The most sensitive end point of toxicity should continue to be used for establishing the reference dose.

Classification of Evidence of Carcinogenicity

EPA's narrative descriptions of the evidence of carcinogenic hazards are appropriate, but a simple classification scheme is also needed for decision-making purposes. The current EPA classification scheme does not capture information regarding the relevance to humans of animal data, any limitations regarding the applicability of observations, or any limitations regarding the range of carcinogenicity outside the range of observation. The current system might thus understate or overstate the degree of hazard for some substances.

- EPA should provide comprehensive narrative statements regarding the hazards posed by carcinogens, to include qualitative descriptions of both: 1) the strength of evidence about the risks of a substance; and 2) the relevance to humans of the animal models and results and of the conditions of exposure (route, dose, timing, duration, etc.) under which carcinogenicity was observed to the conditions under which people are likely to be exposed environmentally. EPA should develop a simple classification scheme that incorporates both these elements. A similar scheme to that set forth in Table 7-1 is recommended. The agency should seek international agreement on a classification system.

Potency Estimates

EPA uses estimates of a chemical's potency, derived from the slope of the dose-response curve, as a single value in the risk-assessment process.

- EPA should continue to use potency estimates—i.e., unit cancer risk—to estimate an upper bound on the probability of developing cancer due to lifetime exposure to one unit of a carcinogen. However, uncertainty about the potency estimate should be described as recommended in Chapter 9.

Although EPA routinely cites available human evidence, it does not always rigorously compare the quantitative risk-assessment model based on rodent data with available information on molecular mechanisms of carcinogenesis or with available human evidence from epidemiological studies.

- Because the validity of the overall risk-assessment model depends on how well it predicts health effects in the human population, EPA should acquire additional expertise in areas germane to molecular and mechanistic toxicology. In addition, EPA should also acquire additional epidemiological data to assess the validity of its estimates of risk. These data might be acquired in part by formalizing a relationship with the National Institute for Occupational Safety and Health to facilitate access to data from occupational exposures.

8
DATA NEEDS

This chapter discusses the quantity, quality, and availability of data needed for conducting an adequate risk assessment in the context of the Clean Air Act Amendments of 1990 (CAAA-90). It begins by discussing the need for a priority-setting process, and the need for an iterative data-collection process. It then indicates the proper prioritization for data collection and the availability of data in each of the key risk-assessment steps. It concludes with a discussion of how data should be managed.

CONTEXT OF DATA NEEDS

Most would agree that, given the best available model, additional relevant data will lead to a more accurate and precise risk assessment. The quality of the data is critical, no matter how excellent the model chosen, to avoid the classic "garbage in, garbage out" problem. In the gathering of data, tradeoffs must often be made among data that are necessary, data that are desirable, and data that are affordable. Desirability must be defined in the context of the risk-management goals to be achieved, which might be the development of regulations, the setting of standards, or the screening of chemicals to set priorities.

The more precisely the risk manager frames the questions to be addressed by the risk assessment at the outset, the less ambiguity there will be as to what data are required to answer the questions, the less need for judgment in data-gathering, and the lower the likelihood that inappropriate or insufficient data will be gathered. As a corollary, public input into the framing of goals and questions can help to avoid public criticism and distrust of the process of risk assessment, including the gathering of exposure and toxicity data. Public confidence that risk managers are addressing real concerns, as opposed to going through a process perfunctorily, is critical to the future of risk assessment as an activity capable of improving the quality of life. Risk managers need to articulate clearly from the beginning who is to be protected from what, when and where, and at what cost (including how much effort and funds are to be expended to collect appropriate data), so that risk assessors can provide relevant information.

IMPLICATIONS FOR PRIORITY-SETTING

It is not necessary, nor would it be cost-effective, to collect all the data needed for a complete health-hazard assessment on all the 189 chemicals (or mixtures) listed in CAAA-90. It is important, however, that the entire list be examined to identify chemicals that are potentially hazardous and that the later full-scale evaluation of each chemical selected for further scrutiny proceed as effectively as possible. An overall strategy is essential for setting priorities among the steps in the information-gathering process and for determining the extent of assessment needed.

Because risk is a function of exposure, as well as toxicity, determining both that a chemical is of low toxicity to all humans and that all humans have only small exposures to it would lead to an overall low priority for a full-scale risk assessment. Obviously, assigning a high priority to both would lead to an overall high priority for such assessment and argue for collection of a complete data set in all categories of exposure and toxicity. There will be various intermediate levels between low and high overall priority.

In the absence of pertinent human data, toxicological evaluation should begin with the simplest, most rapid, and most economical tests and proceed to more complex, time-consuming, and more expensive tests only as warranted by the initial steps. Similarly, emission, transport, and exposure data might be used to rank chemicals for testing, from those with relatively large exposure potential down to those with a very low likelihood of significant exposure, either for the population at large or for any substantial subset of the population. What is "substantial" in this context will of course depend on concurrent assessments of toxicity. Ordering can then be based on an evaluation of a relatively modest or limited data set.

To assess whether there is a potential for exposure, and to gauge the magnitude and duration of exposure, one needs to know:

1. Is the chemical emitted into the air?
2. Is the chemical stable enough to be transported from its source to a population?

If the chemical is not emitted or is so unstable that it breaks down into innocuous products before reaching a population, no further data need be collected and further risk assessment is not warranted. But if it is emitted and can be transported to a population, one needs to ask:

3. Who is exposed, to how much, and for how long?
4. What is the relationship between exposure (dose) and response (effect) for humans and for animals?

In an iterative data-collection process, one works through data related to questions 1-4, first collecting the most critical data within each category, then judging needs for more data within that category before moving to the next category. The process is iterative until sufficient information is gathered to draw a conclusion—e.g., on a potential threat to public health.

Section 112 of the Clean Air Act mandates that EPA consider the hazards and possible regulation of 189 specified chemicals. Considering both the effort required to carry out

complete risk assessments and the resources of the agency, it is unlikely that that can be accomplished within the time constraints of the act. Consequently, in the spirit of the act and in the interest of the public welfare, it is critical that EPA assign priorities to the chemicals listed. These priorities should be based first on their potential impact on human health and welfare.

Some of the 189 chemicals appear to present major problems because of their variety of sources, large exposures, or high potency. Other chemicals present simpler problems—e.g., some have relatively few sources, some have lower potential for human exposures, and some have very low potency. It is an inefficient use of resources to invest huge amounts of money and time in research and analysis to determine factors already known to be inconsequential for final risk assessment or to confirm credible estimates on which consensus can easily be obtained. Therefore, EPA should do preliminary analyses (screenings) on all listed compounds to ascertain which chemicals merit detailed risk-assessment efforts and which do not merit such work. These preliminary analyses should be reviewed by an independent board to ensure the validity of the resulting priorities for full-scale assessments. Priorities should be continually re-evaluated and changed as appropriate in response to new data. The task of setting priorities and keeping them up to date is not trivial and should be specifically included, with adequate resources, in EPA's evolving program plan to implement CAAA-90. The iterative data-collection process can then help in setting priorities for ranking needed studies to avoid the accumulation of a surfeit of data, which would result in misuse of funds and waste of time.

DATA NEEDED FOR RISK ASSESSMENT

The following sections discuss the priority-setting and availability of data for each of the key data-processing steps in risk assessment: emissions, environmental fate and transport, exposure, and toxicity. The final section summarizes the data priorities in each of these areas, and indicates how this data can be used for overall priority-setting for data collection.

EMISSIONS

Knowledge of emissions of a chemical into the air—specifically, the quantity emitted per unit of time (flux) from each place where it is made, stored, used, or disposed of plus its physical and chemical form—is fundamental to characterizing the magnitude of expected exposure to the chemical.

PRIORITIES FOR COLLECTING DATA

The specific methods for characterizing emissions are described and evaluated in Chapter 7.

On the basis of this analysis, an iterative data-collecting process for emission characterization might proceed roughly as follows:

1. Plant-specific material balance
2. Industry-wide emission factors
3. Plant-specific emission factors
4. Facility measurements, including flux determinations.

Data quality is critical, because of the wide variety of emission-estimation techniques and the many types of facilities emitting hazardous air pollutants. EPA often uses whatever data are available at the time of decision-making and has not published guidelines or standards for the quality of emission data to be used in its risk assessments.

Because the emission-characterization database is extremely important for priority-setting, EPA should review the emission estimates submitted to ensure that they meet reasonable quality standards and that emission estimates from all sources within a site are submitted.

DATA AVAILABILITY

EPA plans to use emission information that is available in the Toxic Release Inventory (TRI) database as required by Title III of the Superfund Amendments and Recovery Act (SARA). The information available in this database is shown in the table provided by EPA to the committee in Appendix A. The TRI database includes information on annual emissions, facility location, and categorization of emissions as fugitive, point source, or both.

These data have two serious limitations for any use in risk assessment. First, the database does not include emissions from all operations at a facility; for example, transfer operations are not reported. Second, the database does not include emissions of less than 10 tons/year, nor does it have the locations of emission points or the frequency of emissions. Some information is available in emission inventory databases that are required by state implementation plans (SIPs) that states are required to submit to EPA to indicate how they plan to control emissions relative to CAAA-90, but that information is not necessarily well characterized. For example, emissions of volatile organic chemicals (VOCs) might be listed as a total, instead of as emissions of separate chemicals; but risk assessments should generally be done for separate chemicals, rather than for classes of chemicals.

A study by Amoco and EPA (1992) gives an example of the differences between estimated or calculated emissions (such as those listed in the TRI database) and emissions determined via direct measurement. This study found that the "existing estimates of environmental releases were not adequate for making a chemical-specific, multi-media, facility wide assessment." The report identified several specific problems in using the TRI database to conduct an in-depth evaluation of a facility:

- Lack of chemical characterization data.

TABLE 8-1. List of Section 112 pollutants not in toxic release inventory data base.

2,2,4-Trimethyl pentane
Acetophenone
Caprolactan
Dichlorodiphenyldichloroethylene (DDE)
Dimethyl formanide
Fine mineral fibers
1 texamethylene-t,t,-diisocyanate
Hexane
Isophorone
Phosphine
Polycylic organic matter
Sulfur dioxide, anhydrous
TCDD
Triethylamine

PRIORITIES FOR COLLECTING DATA

In the proposed iterative data-collection process described at the beginning of this chapter, data on environmental fate and transport would be acquired in roughly the following order:

1. Physical properties.
2. Physicochemical properties of environment.
3. Chemical properties or reactivity.
4. Rates of potential removal processes.

Once that information is available, a model calculation of expected concentrations in nearby air is relatively straightforward. If the information is not available, it must be obtained or assumed.

Data Availability

Data on emissions and physical properties are generally available or can be estimated (Lyman et al., 1982). For chemical properties and reactivity, they are available for some environmental reactions, but not all. In the case of physicochemical properties, the environment data are generally available at most locations in the United States. Information on the rates of potential removal processes are more difficult and costly to obtain.

Careful evaluation of data is necessary. For example, published vapor pressures of organic chemicals of moderate to low volatility determined under laboratory conditions can be seriously inaccurate and misleading. For all chemicals, vapor-phase reaction rate constants, when extrapolated from the laboratory to outdoor ambient air, can be seriously in error. The literature is not always for purposes of risk assessment.

EXPOSURE

Accurate exposure data are crucial to valid risk assessment. For example, exposure data

- Difficulty of measuring and characterizing small sources.
- Use of estimated, rather than actual, data.
- Lack of identification of new sources leading to underestimation.
- Overestimation of some sources because of use of standardized industry-wide emission factors.
- No requirement that all chemicals be reported in the TRI database (e.g., only 9% of total hydrocarbons were required to be reported).
- Exclusion of some activities and emissions from record-keeping requirements (e.g., barge loading, which accounted for about 20% of benzene emissions).
- Lack of data in TRI on location of nearby populations and ecosystems.

EPA should develop a mechanism to gather the information just listed in a consistent fashion. This mechanism could include changes in Title III of SARA, which requires the TRI reporting requirement or development of information for Title I or V of CAAA-90. Although development of emission characterization databases for all of the 189 chemicals might initially seem to be a major task, CAAA-90 requires states to develop more detailed emission inventories by November 1992 and to update them. Most facilities are then required to estimate their emissions on a point basis to satisfy state requirements for emission inventories. Much of this information is also required for permit purposes.

Even simple changes, such as modifying the SARA Title III requirements to include all 189 hazardous air pollutants on the list, would help. Sixteen of the 189 compounds in CAAA-90 Title III are not on the TRI list (see Table 8-1). In addition, the TRI database includes only sources that have 10 or more full-time employees and that manufacture, process, or use specified chemicals above a certain production rate. That restriction excludes smaller sources within the manufacturing sector for which risk assessments must be conducted under the Title III requirements. Instituting an emission threshold relative to the Title III requirements (e.g., 10 tpy for single compound; 25 tpy for multiple compounds) might be more appropriate for gathering information for risk-assessment purposes.

For evaluation of VOCs, many of which are on the list of 189 compounds under Title III, emission estimates developed for other regulatory purposes (such as the ozone provisions of CAAA-90) can be used. However, these data are frequently not speciated in terms of the chemical composition of the VOCs. In addition, the reporting of VOC emission information is required only in nonattainment areas, so this information may not always be available.

Environmental Fate and Transport

Emitted pollutants can move within and between environmental media and be converted to different forms. A thorough understanding of what happens to a chemical in the environment forms part of the basis for estimating human exposure and hence determining risk.

must match up temporally with the health end points of concern. Key issues in the evaluation of exposure are

- The end points of interest (e.g., acute vs. chronic toxicity).
- The populations at risk (i.e., the general population and defined subpopulations with potentially increased risks).
- The routes of exposure (e.g., air, diet, or skin).
- The duration (e.g., lifetime, annual, or instantaneous).
- The nature and degree of simultaneous toxicant exposures.

Rarely are all those issues resolved by the exposure data available for a risk assessment. Efforts to collect the data should focus on the minimum needed to meet the goals of the assessment in its risk-management context.

PRIORITIES FOR COLLECTING DATA

In the proposed iterative data-collection process, the order of data collection might be as follows:

1. *Ambient-air monitoring.* Most commonly, ambient-air monitoring produces interval concentrations in samples averaged over a fixed time, such as 8 hr or 24 hr at fixed sampling stations. The number of stations, their times of operation, and their locations relative to known emission sources and populations at risk must be known, as well as concentration averages, variances or ranges (to estimate uncertainty), and a description of the methods used, including potential error. The time interval of ambient-air monitoring should be commensurate with the time needed to elicit the physiological effects of concern.
2. *Targeted fixed-point monitoring data.* These data are often generated from samples placed near sources of high-volume emissions (i.e., "hot spots") or in response to some real or perceived public-health need. They should be accompanied by the same information as for ambient-air monitoring. Targeted monitoring is often more useful than monitoring at pre-existing sampling stations if it can focus on higher concentrations of a pollutant, a population at greater risk, or both.
3. *Peak-concentration data.* Either ambient-air or targeted monitoring can miss peak concentrations, because the sampling interval is so long as to "average out" all peaks and valleys in the sampled air mass. Sampling with instantaneous analyzers (e.g., spectrophotometers) or interval analyzers that can accept a sample of short duration is needed to define peaks. That might be of special importance for a toxicant released intermittently.
4. *Personal monitoring.* Concentration data from personal monitors are often more useful for risk assessment, because they show the exposure of individual subjects and can be used to relate activity patterns to exposure. If enough subjects are selected for monitoring, a population exposure can be constructed. Such information is not yet generally available, except for a

few toxicants, because of the time and expense of a comprehensive study. This in turn is primarily due to a lack of low-cost, portable sampling devices for most chemicals. Active samplers may provide more information directly for risk assessment than passive samplers for personal monitoring, because pollutant concentrations (and thus the dose) can be estimated more directly with active sampling. Passive samplers do not provide specific concentrations; however, they are far less costly and bulky than active samplers. They are useful in screening (i.e., to determine whether exposure has occurred). Research to correlate the concentrations detected by passive samplers with exposure and dose would further enhance their potential.

5. *Biological markers.* If a toxicant produces a metabolite, enzyme alteration, or other signal that exposure has occurred and so leads to a high correlation between that marker and degree of exposure, such information can reduce the uncertainty in a predicted risk and could be useful for risk assessment. In one respect, this would be the best exposure information, because it would show that the toxicant has been absorbed and has already had some biological effect (NRC, 1987); but it makes single-source exposure assessment difficult, because it reveals total uptake across all routes of exposure. Unless biologic-marker data are checked against external exposure data, they cannot be used to determine dose. Validation of the correlation between an external concentration and the magnitude of a biological marker in experimental animals can be helpful, but one is left with the difficulty of extrapolating to humans, who may not respond in the same quantitative way as experimental animals. In some cases, markers in humans can be established in occupational settings.

DATA AVAILABILITY

Some of the 189 chemicals on the Clean Air Act Amendments list have relatively abundant data on concentrations; some have virtually none. When concentration data are available, they are more likely to be from ambient-air monitoring or, at best, targeted fixed-point monitoring. For only some of the compounds are sufficient exposure data available for preliminary evaluation of relative priority for more detailed risk assessment (see Appendix A). That is a major problem that can be solved only by a much more extensive state or federal monitoring program. Some states, such as California, are moving rapidly in developing a hazardous air-pollutant monitoring program. Coordination between states and with federal agencies is necessary to keep scarce resources from being wasted in duplicative efforts.

Collection of new exposure data on humans is limited by current methods of monitoring individual exposures (which are often expensive, often of low accuracy or precision, and often nonquantitative or lacking in the ability to determine the source of exposure) and by methods of obtaining information on human behavior that might affect uptake or exposures. In addition, no reference database is available for comparing new data, that is, for determining whether new data represent exposure outside the general norm or are within the realm of acceptability defined by prior studies. Furthermore, when exposure data are gathered, they should be probability-based to allow inferences to the population and estimation of the tails of the distribution of exposures.

TOXICITY

A full assessment of the inherent toxicity of an agent requires some combination of structure-activity analyses, in vitro or whole-animal short-term tests, chronic or long-term animal bioassays, human biomonitoring, clinical studies, and epidemiological investigations (NRC, 1984, 1991c,d). A complete hazard identification might entail review of information in all those categories before a determination that a quantitative risk assessment of the agent is warranted (Bailar et al., 1993).

Estimation of dose-effect relationships requires data on the effects of a wide range of doses, on factors that influence the dose delivered to critical target cells by given magnitudes and patterns of exposure (e.g., uptake, anatomic distribution, metabolism, and excretion) (NRC, 1987), on the shapes and slopes of pertinent dose-effect curves, on the relevant mechanisms of effects (NRC, 1991c), and on the extent to which the response to an agent can vary with species, sex, age, previous exposure, health status, exposure to extraneous agents, and other variables (NRC, 1988a).

PRIORITIES FOR COLLECTING DATA

Strategies to fill data gaps in toxicity assessment are best developed case by case, but the following priority-setting of the major types of toxicological data that may be used are listed below. In the suggested iterative data-collection process, the toxicity data listed in the first three categories below (i.e., generic and acute toxicity, acute mammalian lethality) should be collected on every chemical as a starting point, and other, more expensive, data should be collected only on chemicals that give cause for concern based on the data in those categories.

1. Generic toxicity data (structure-activity relationships and results of other correlational analyses).
2. Data on acute toxicity (on lethality in microorganisms or effects on mammalian cells in vitro).
3. Acute mammalian lethality data (usually rodent).
4. Toxicokinetics data, phase 1 (on uptake, distribution, retention, and excretion in rodents).
5. Genotoxicity data (results of short-term in vitro tests in microorganisms, *Drosophila*, and mammalian cells).
6. Data on subchronic toxicity (on 14-day or 28-day inhalation toxicity in rodents).
7. Toxicokinetic data, phase 2 (on metabolic pathways and metabolic fate in rodents and other mammalian species, with special attention given to exposure by inhalation).
8. Data on chronic toxicity (on carcinogenicity, neurobehavioral toxicity, reproductive and developmental toxicity, and immunotoxicity in two rodent species of both sexes, with special attention given to the exposure by inhalation).

9. Human toxicity data (clinical, biomonitoring, and epidemiological data).
10. Data on toxic mechanisms, dose-effect relationships, influence of modifying factors (age, sex, and other variables) on susceptibility, and interactive effects of mixtures of chemical and physical agents.

This prioritization is based on the cost and complexity of gathering such data (NRC, 1984). It is generally not possible to plan the collection of clinical and epidemiological data. Toxicological studies conducted clinically in humans are usually planned and implemented under experimental control, but very few are done, because of the attendant hazards. Epidemiological studies are relatively expensive and often produce data that are difficult to interpret as to effects of specific toxic agents. If one were to set data-collection priorities without concern for cost, ethical, or other considerations, the sequence of collection might be

1. Toxicological human data.
2. Clinical data.
3. Epidemiological data.

Data Availability

Availability of requisite data varies widely among the 189 chemicals. On the one hand, some preliminary toxicity data are available on some of the chemicals, or at least can be estimated from structure-activity correlations. On the other hand, the toxicity data are incomplete on almost all 189 chemicals.

The amount of data available is highly variable and depends largely on the existence of uncontrollable chance events. Generally, better data sets exist on individual chemicals that have been used over long periods (vinyl chloride, some solvents, etc.) and on chemicals of wide use (such as pesticides) than on chemicals rarely used or chemicals that are byproducts of other chemicals (e.g., chemicals in automobile exhaust and cigarette smoke). Additional information and analysis on the Integrated Risk Information System (IRIS) used by EPA is provided in Chapter 12. Some of the partial data needed to test models are discussed in Chapter 6.

Overall Priority Setting

The data needed for each step of risk assessment are summarized in rough order of increasing complexity (see Table 8-2). In an iterative data-collection process, if information in the top one or two items of each of the four columns in Table 8-2 does not indicate increased risk potential the priority for full risk assessment should be low. Various combinations of negative information in the first few items of any two of the first three lists (e.g., emissions, environmental fate and transport, exposure) with positive information in the third list might lead to a

TABLE 8-2. Types of data available for risk assessment.

Emissions	Environmental fate and transport	Exposure	Toxicity
1. Material balance	1. Physical properties	1. Ambient fixed-point monitoring	1. Generic toxicity
2. Industry-wide emission factors	2. Physicochemical properties of environment	2. Targeted fixed-point monitoring	2. Acute toxicity (lethality for microorganisms or mammalian cells in vitro)
3. Plant-specific emission factors (EPA protocol)	3. Chemical properties or reactivity	3. Duration and frequency of peak concentrations for populations at risk	3. Acute mammalian lethality (rodent)
4. Facility measurements, including flux determinations	4. Rates of potential removal processes	4. Personnel monitoring for average and maximally exposed people	4. Toxicokinetics, phase 1
		5. Biologic markers	5. Genotoxicity (short-term in vitro tests in microorganisms, Drosophila, or mammalian cells)
			6. Subchronic (13-day or 28-day) inhalation toxicity (rodent)
			7. Toxicokinetics, phase 2
			8. Chronic toxicity: carcinogenicity, neurobehavioral toxicity, reproductive and developmental toxicity, or immunotoxicity
			9. Human toxicity (clinical, biomonitoring, epidemiologic)
			10. Toxic mechanisms and dose-effect relationships

medium priority. Positive information in the early items of two, or perhaps three, of the lists would argue for a high priority. Data for the more complex items of each list would be developed when evidence of potential hazard exceeded an agreed-on "bright line" of concern, i.e., a decision point set either by regulation or programmatic procedures.

Although a full priority scheme probably should be on a continuous scale, several important points to develop a more detailed scheme might appear as follows:

SCREENING RISK ASSESSMENT

Emissions—Items 1 and 2
Environmental fate and transport—Items 1-3
Exposure—Items 1-3
Toxicity—Items 1-3

- If the information for all the above items (or items lower on the list, if available) indicates no potential health concerns, assign "low priority."
- If any information on exposure (emissions, environmental fate and transport, exposure) is positive, assign the chemical "medium priority".
- If any information on exposure is positive (i.e., emission, environmental fate and transport, or exposure measurement), *and* toxicity data are positive, then assign the chemical "high priority"and proceed to the full-scale risk assessment.

FULL RISK ASSESSMENT

Emissions—Items 1-4
Environmental fate and transport—Items 1-5
Exposure—Items 1-5
Toxicity—Items 1-10

- If the information is not positive for the higher-order items in all four lists, assign the chemical to Action Level 2 (more extended time response).
- If the information is positive for the higher-order items in all four lists, assign the chemical to Action Level 1 (short time-frame response).

Reliable positive human evidence will always result in a high priority and the full risk evaluation. Any positive clinical, toxicologic, or epidemiological human data would override a priority based on exposure and animal toxicity data alone and move a given chemical to the stage of full risk assessment.

The detailed nature of the process used to set priorities for full risk assessment needs to be addressed in a coordinated way by federal and state agencies, to ensure the best use of limited

resources for this programmatic step. There might be, for example, a numerical weighting or scoring approach based on data in the four categories of emissions, environmental fate and transport, exposure, and toxicological data. EPA should consider convening a panel of experts to develop a priority-setting process and the requisite accompanying iterative approach to data collection.

Data Management

More attention needs to be paid to data management to ensure that vital data gaps are filled, that data used in risk assessments are of the best possible quality, and that relevant information (such as negative epidemiological information) is not overlooked. The lack of a consistent data-collection scheme makes data analysis, and thus effective risk assessment, inconsistent and unreliable for risk-management purposes.

For example, risk assessment often requires that the assessor decide whether to set aside information from old studies when newer, supposedly better information is available. The ultimate desire is for credibility; therefore, it is important to use information that is widely acknowledged as the best representation of reality. If the results of a new study contradict information from an old study and if there is only a small difference in the "bottom-line" estimate of human health risk, then both should be used, and the error bounds of the current risk assessment should be revised. However, if the studies lead to quite different conclusions, use of both might be feasible. For example, some animal evidence might show a major health hazard while there may also be weak, negative, or equivocal animal studies. Such conflicting data should be carefully reviewed in the risk-assessment document, with detailed study of possible reasons for the discrepancy. When no reconciliation of results seems feasible, the committee recommends that the voice of prudence be heard and that the risk assessment be either based on the higher ultimate risk estimate or delayed (as was done in part on formaldehyde) until additional studies can be completed.

Findings and Recommendations

The committee's findings and recommendations follow.

Insufficient Data for Risk Assessment

EPA does not have sufficient data to assess fully the health risks of the 189 chemicals in Title III within the time permitted by the Clean Air Act Amendments of 1990.

- EPA should screen the 189 chemicals for priorities for the assessment of health risks,

identify the data gaps, and develop incentives to expedite generation of the needed data by other public agencies (such as the National Toxicology Program, the Agency for Toxic Substances and Disease Registry, and state agencies) and by other organizations (industry, academia, etc.).

Need for Data-Gathering Guidelines

EPA has not defined the guidelines or process to be used for determining the types, quantities, and quality of data that are needed for conducting risk assessments for facilities emitting one or more of the 189 chemicals.

- EPA should develop an iterative approach to gathering and evaluating data in the categories of emission, transport and fate, exposure, and toxicology for use in both screening and full risk assessment. The data-gathering and data-evaluation process should be set forth by EPA in guidelines for use by those who conduct data-gathering activities. To develop these guidelines, EPA should convene a panel of experts to develop a priority-setting scheme that uses a numerical weighting or scoring approach.

Inadequacy of Emission and Exposure Data

EPA has often relied on non-site-specific emission and exposure data. These data are often not sufficient to assess the risk to individuals and the affected population at large.

- EPA should expand its efforts to gather emission and exposure data to personal monitoring and site-specific monitoring.

Inadequacy of TRI Database as a Source of Emission Data for Risk-Assessment Purposes

The SARA 313 Toxic Release Inventory data and other readily available data used by EPA for emission characterization may be adequate for screening purposes but are not adequate for developing detailed risk assessments for specific facilities. Present processes of gathering emission data do not yield information appropriate for all risk-assessment purposes under the Clean Air Act Amendments.

- EPA should modify its data-gathering activities related to emissions to ensure that it has or will acquire the data needed to conduct screening and full risk assessments, especially of the 189 chemicals listed in CAAA-90.

Lack of Adequate Natural Background-Exposure Database

EPA does not have an adequate database on natural background exposures to the 189 air pollutants against which to evaluate total human exposure data from facilities producing or using these substances.

- EPA should develop an ambient-outdoor-exposure database on the 189 listed hazardous air pollutants.

Inadequate Explanation of Analytical Techniques

EPA does not always explain adequately the analytical and measurement methods it uses for estimating ambient outdoor exposures.

- EPA should collate and explain the analytical and measurement methods it uses for ambient outdoor exposures, including the errors, precision, accuracy, detection limits, etc., of all methods that it uses for risk-assessment purposes.

Need for System of Data Management for Risk Assessment

EPA needs more adequate mechanisms to compile and maintain databases for use in health-risk screening and assessment.

- EPA should review its data-management systems and improve them as needed to ensure that the quality and quantity of the data are routinely updated and that the data are sufficiently accessible for risk screening and risk assessment. Its responsibilities under CAAA-90 should be prominent in this review and revision.

9

UNCERTAINTY

The need to confront uncertainty in risk assessment has changed little since the 1983 NRC report *Risk Assessment in the Federal Government*. That report found that:

> The dominant analytic difficulty [in decision-making based on risk assessments] is pervasive uncertainty. . . . there is often great uncertainty in estimates or the types, probability, and magnitude of health effects associated with a chemical agent of the economic effects of a proposed regulatory action, and of the extent of current and possible future human exposures. These problems have no immediate solutions, given the many gaps in our understanding of the causal mechanisms of carcinogenesis and other health effects and in our ability to ascertain the nature or extent of the effects associated with specific exposures.

Those gaps in our knowledge remain, and yield only with difficulty to new scientific findings. But a powerful solution exists to some of the difficulties caused by the gaps: the systematic analysis of the sources, nature, and implications of the uncertainties they create.

CONTEXT OF UNCERTAINTY ANALYSIS

EPA decision-makers have long recognized the usefulness of uncertainty analysis. As indicated by former EPA Administrator William Ruckelshaus (1984):

> First, we must insist on risk calculations being expressed as distributions of estimates and not as magic numbers that can be manipulated without regard to what they really mean. We must try to display more realistic estimates of risk to show a range of probabilities. To help do this, we need new tools for quantifying and ordering sources of uncertainty and for putting them into perspective.

Ten years later, however, EPA has made little headway in replacing a risk-assessment "culture" based on "magic numbers" with one based on information about the range of risk values consistent with our current knowledge and lack thereof.

As we discuss in more depth in Chapter 5, EPA has been skeptical about the usefulness of uncertainty analysis. For example, in its guidance to those conducting risk assessments for Suprfund sites (EPA, 1991f), the agency concludes that quantitative uncertainty assessment is usually not practical or necessary for site risk assessments. The same guidance questions the value and accuracy of assessments of the uncertainty, suggesting that such analyses are too data-intensive and "can lead one into a false sense of certainty."

In direct contrast, the committee believes that uncertainty analysis is the only way to combat the "false sense of certainty," which is *caused* by a refusal to acknowledge and (attempt to) quantify the uncertainty in risk predictions.

This chapter first discusses some of the tools that can be used to quantify uncertainty. The remaining sections discuss specific concerns about EPA's current practices, suggest alternatives, and present the committee's recommendations about how EPA should handle uncertainty analysis in the future.

NATURE OF UNCERTAINTY

Uncertainty can be defined as a lack of precise knowledge as to what the truth is, whether qualitative or quantitative. That lack of knowledge creates an intellectual problem— that we do not know what the "scientific truth" is; and a practical problem—we need to determine how to assess and deal with risk in light of that uncertainty. This chapter focuses on the practical problem, which the 1983 report did not shed much light on and which EPA has only recently begun to address in any specific way. This chapter takes the view that uncertainty is always with us and that it is crucial to learn how to conduct risk assessment in the face of it. Scientific truth is always somewhat uncertain and is subject to revision as new understanding develops, but the uncertainty in quantitative health risk assessment might be uniquely large, relative to other science-policy areas, and it requires special attention by risk analysts. These analysts need to allow questions such as: What should we do in the face of uncertainty? How should it be identified and managed in a risk assessment? How should an understanding of uncertainty be forwarded to risk managers, and to the public? EPA has recognized the need for more and better uncertainty assessment (see EPA memorandum in Appendix B), and other investigators have begun to make substantial progress with the difficult computations that are often required (Monte Carlo methods, etc.). However, it appears that these changes have not yet affected the day-to-day work of EPA.

Some scientists, mirroring the concerns expressed by EPA, are reluctant to quantify uncertainty. There is concern that uncertainty analysis could reduce confidence in a risk assessment. However, that attitude toward uncertainty may be misguided. The very heart of risk assessment is the responsibility to use whatever information is at hand or can be generated to produce a number, a range, a probability distribution—whatever expresses best the present state of knowledge about the effects of some hazard in some specified setting. Simply to ignore the uncertainty in any process is almost sure to leave critical parts of the process incompletely examined, and hence to increase the probability of generating a risk estimate that is incorrect, incomplete, or misleading.

For example, past analyses of the uncertainty about the carcinogenic potency of saccharin showed that potency estimates could vary by a factor as large as 10^{10}. However, this example is not representative of the ranges in potency estimates when appropriate models are compared. Potency estimates can vary by a factor of 10^{10} only if one allows the choice of some models that are generally recognized as having no biological plausibility and only if one uses those

models for a very large extrapolation from high to low doses. The judicious application of concepts of plausibility and parsimony can eliminate some clearly inappropriate models and leave a large but perhaps a less daunting range of uncertainties. What is important, in this context of enormous uncertainty, is not the best estimate or even the ends of this 10^{10}-fold range, but the best-informed estimate of the likelihood that the true value is in a region where one rather than or another remedial action (or none) is appropriate. Is there a small chance that the true risk is as large as 10^{-2}, and what would be the risk-management implications of this very small probability of very large harm? Questions such as these are what uncertainty analysis is largely about. Improvements in the understanding of methods for uncertainty analysis—as well as advances in toxicology, pharmacokinetics, and exposure assessment—now allow uncertainty analysis to provide a much more accurate, and perhaps less daunting, picture of what we know and do not know than in the past.

TAXONOMIES

Before discussing the practical applications of uncertainty analysis, it may be best to step back and discuss it as an intellectual endeavor. The problem of uncertainty in risk assessment is large, complex, and nearly intractable, unless it is divided into smaller and more manageable topics. One way to do so, as seen in Table 9-1 (Bogen, 1990a), is to classify sources of uncertainty according to the step in the risk assessment process in which they occur. A more abstract and generalized approach preferred by some scientists is to partition all uncertainties into the three categories of bias, randomness, and true variability. This method of classifying uncertainty is used by some research methodologists, because it provides a complete partition of types of uncertainty, and it might be more productive intellectually: bias is almost entirely a product of study design and performance; randomness a problem of sample size and measurement imprecision; and variability a matter for study by risk assessors but for resolution in risk management (see Chapter 10).

However, a third approach to categorizing uncertainty may be more practical than this scheme, and yet less peculiar to environmental risk assessment than the taxonomy in Table 9-1.

This third approach, a version of which can be found in EPA's new exposure guidelines (EPA, 1992a) and in the general literature on risk assessment uncertainty (Finkel, 1990; Morgan and Henrion, 1990), is adopted here to facilitate communication and understanding in light of present EPA practice. Although the committee makes no formal recommendation on which taxonomy to use, EPA staff might want to consider the alternative classification above (bias, randomness, and variability) to supplement their current approach in future documents. Our preferred taxonomy consists of:

- *Parameter uncertainty*. Uncertainties in parameter estimates stem from a variety of sources. Some uncertainties arise from measurement errors; these in turn can involve random errors in analytic devices (e.g., the imprecision of continuous monitors that measure stack

TABLE 9-1. Some generic sources of uncertainty in risk assessment.

I. HAZARD IDENTIFICATION

Unidentified hazards

Definition of incidence of an outcome in a given study (positive-negative association of incidence with exposure)

Different study results

Different study qualities
- conduct
- definition of control population
- physical-chemical similarity of chemical studied to that of concern

Different study types
- prospective, case-control, bioassay, in vivo screen, in vitro screen
- test species, strain, sex, system
- exposure route, duration

Extrapolation of available evidence to target human population

II. DOSE-RESPONSE ASSESSMENT

Extrapolation of tested doses to human doses

Definition of "positive responses" in a given study
- independent vs. joint events
- continuous vs. dichotomous input response data

Parameter estimation

Different dose-response sets
- results
- qualities
- types

Model selection for low-dose risk extrapolation
- low-dose functional behavior of dose-response relationship (threshold, sublinear, linear, supralinear, flexible)
- role of time (dose frequency, rate, duration; age at exposure; fraction of lifetime exposed)
- pharmacokinetic model of effective dose as a function of applied dose
- impact of competing risks

TABLE 9-1. continued

III. EXPOSURE ASSESSMENT

Contamination-scenario characterization (production, distribution, domestic and industrial storage and use, disposal, environmental transport, transformation and decay, geographic bounds, temporal bounds)
- environmental-fate model selection (structural error)
- parameter estimation error
- field measurement error

Exposure-scenario characterization
- exposure-route identification (dermal, respiratory, dietary)
- exposure-dynamics model (absorption, intake processes)

Target-population identification
- potentially exposed populations
- population stability over time

Integrated exposure profile

IV. RISK CHARACTERIZATION

Component uncertainties
- hazard identification
- dose-response assessment
- exposure assessment

Source: Bogen, 1990a.

emissions) or systematic biases (e.g., measuring inhalation from indoor ambient air without considering the effect of volatilization of contaminants from hot water used in showering). A second type of parameter uncertainty arises when generic or surrogate data are used instead of analyzing the desired parameter directly (e.g., the use of standard emission factors for industrialized processes). Other potential sources of error in estimates of parameters are misclassification (e.g., incorrect assignment of exposures of subjects in historical epidemiological studies due to faulty or ambiguous information), random sampling error (e.g., estimation of risk to laboratory animals or exposed workers from outcomes observed in only a small sample), and nonrepresentativeness (e.g., developing emission factors for dry cleaners based on a sample that included predominantly "dirty" plants due to some quirk in the study design).[1]

[1]Although variability in a risk-assessment parameter across different individuals is itself a type of uncertainty and is the subject of the following chapter, it is possible that new parameters might be incorporated into a risk assessment to model that variability (e.g., a parameter for the standard deviation of the amount of air that a random person breathes each day) and that those parameters themselves

• *Model uncertainty*. These uncertainties arise because of gaps in the scientific theory that is required to make predictions on the basis of causal inferences. For example, the central controversy over the validity of the linear, no-threshold model for carcinogen dose-response is an argument over model uncertainty. Common types of model uncertainties include relationship errors (e.g., incorrectly inferring the basis for correlations between chemical structure and biologic activity) and errors introduced by oversimplified representations of reality (e.g., representing a three-dimensional aquifer with a two-dimensional mathematical model). Moreover, any model can be incomplete if it excludes one or more relevant variables (e.g., relating asbestos to lung cancer without considering the effect of smoking on both those exposed to asbestos and those unexposed), uses surrogate variables for ones that cannot be measured (e.g., using wind speed at the nearest airport as a proxy for wind speed at the facility site), or fails to account for correlations that cause seemingly unrelated events to occur much more frequently than would be expected by chance (e.g., two separate components of a nuclear plant are both missing a particular washer because the same newly hired assembler put both of them together). Another example of model uncertainty concerns the extent of aggregation used in the model. For example, to fit data on the exhalation of volatile compounds adequately in physiologically based pharmacokinetic (PBPK) models, it is sometimes necessary to break up the fat compartment into separate compartments reflecting subcutaneous and abdominal fat (Fiserova-Bergerova, 1992). In the absence of enough data to indicate the inadequacy of using a single aggregated variable (total body fat), the modeler might construct an unreliable model. The uncertainty in risk that results from uncertainty about models might be as high as a factor of 1,000 or even greater, even if the same data are used to determine the results from each. This can occur, for example, when the analyst must choose between a linear multistage model and a threshold model for cancer dose-response relations.

Problems with EPA's Current Approach to Uncertainty

EPA's current practice on uncertainty is described elsewhere in this report, especially in Chapter 5, as part of the risk-characterization process. Overall, EPA tends at best to take a qualitative approach to uncertainty analysis, and one that emphasizes model uncertainty rather than parameter uncertainties. The uncertainties in the models and the assumptions made are listed (or perhaps described in a narrative way) in each step of the process; these are then presented in a nonquantitative statement to the decision-maker.

Quantitative uncertainty analysis is not well explored at EPA. There is little internal guidance for EPA staff about how to evaluate and express uncertainty. One useful exception is the analysis conducted for the National Emission Standards for Hazardous Air Pollutants (NESHAPS) radionuclides document (described in Chapter 5), which provides a good initial

---

might be uncertain (see "uncertainty and variability" section in Chapter 11).

example of how uncertainty analysis could be conducted for the exposure portion of risk assessment. Other EPA efforts, however, have been primarily qualitative, rather than quantitative. When uncertainty is analyzed at EPA, the analysis tends to be piecemeal and highly focused on the sensitivity of the assessment to the accuracy of a few specified assumptions, rather than a full exploration of the process from data collection to final risk assessment, and the results are not used in a systematic fashion to help decision-makers.

The major difficulty with EPA's current approach is that it does not supplant or supplement artificially precise single estimates of risk ("point estimates") with ranges of values or quantitative descriptions of uncertainty, and that it often lacks even qualitative statements of uncertainty. This obscures the uncertainties inherent in risk estimation (Paustenbach, 1989; Finkel, 1990), although the uncertainties themselves do not go away. Risk assessments that do not include sufficient attention to uncertainty are vulnerable to four common and potentially serious pitfalls (adapted from Finkel, 1990):

(1) They do not allow for optimal weighing of the probabilities and consequences of error for policy-makers so that informed risk-management decisions can be made. An adequate risk characterization will clarify the extent of uncertainty in the estimates so that better-informed choices can be made.

(2) They do not permit a reliable comparison of alternative decisions, so that appropriate priorities can be established by policy-makers comparing several different risks.

(3) They fail to communicate to decision-makers and the public the range of control options that would be compatible with different assessments of the true state of nature. This makes informed dialogue between assessors and stakeholders less likely, and can cause erosion of credibility as stakeholders react to the overconfidence inherent in risk assessments that produce only point estimates.

(4) They preclude the opportunity for identifying research initiatives that might reduce uncertainty and thereby reduce the probability or the impact of being caught by surprise.

Perhaps most fundamentally, without uncertainty analysis it can be quite difficult to determine the conservatism of an estimate. In an ideal risk assessment, a complete uncertainty analysis would provide a risk manager with the ability to estimate risk for each person in a given population in both actual and projected scenarios of exposures; it would also estimate the uncertainty in each prediction in quantitative, probabilistic terms. But even a less exhaustive treatment of uncertainty will serve a very important purpose: it can reveal whether the point estimate used to summarize the uncertain risk is "conservative," and if so, to what extent. Although the choice of the "level of conservatism" is a risk-management prerogative, managers might be operating in the dark about how "conservative" these choices are if the uncertainty (and hence the degree to which the risk estimate used may fall above or below the true value) is ignored or assumed, rather than calculated.

Some Alternatives to EPA's Approach

A useful alternative to EPA's current approach is to set as a goal a quantitative assessment of uncertainty. Table 9-2, from Resources for the Future's Center for Risk Management, suggests a sequence of steps that the agency could follow to generate a quantitative uncertainty estimate. To determine the uncertainty in the estimate of risk associated with a source probably requires an understanding of the uncertainty in each of the elements shown in Table 9-3. The following pages describe more fully the development of probabilities and the method of using probabilities as inputs into uncertainty analysis models.

Probability Distributions

A probability density function (PDF) describes the uncertainty, encompassing objective or subjective probability, or both, over all possible values of risk. When the PDF is presented as a smooth curve, the area under the curve between any two points is the probability that the true value lies between the two points. A cumulative distribution function (CDF), which is the integral or sum of the PDF up to each point, shows the probability that a variable is equal to or less than each of the possible values it can take on. These distributions can sometimes be estimated empirically with statistical techniques that can analyze large sets of data adequately. Sometimes, especially when data are sparse, a normal or lognormal distribution is assumed and its mean and variance (or standard deviation) are estimated from available data. When data are in fact normally distributed over the whole range of possible values, the mean and variance completely characterize the distribution, including the PDF and CDF. Thus, with certain assumptions (such as normality), only a few points might be needed to estimate the whole distribution for a given variable, although more points will both improve the representation of the uncertainty and allow examination of the normality assumption. However, the problem remains that apparently minor deviations in the extreme tails may have major implications for risk assessment (Finkel, 1990). Furthermore, it is important to note that the assumption of normality may be inappropriate.

When data are flawed or not available or when the scientific base is not understood well enough to quantify the probability distributions of all input variables, a surrogate estimate of one or more distributions can be based on analysis of the uncertainty in similar variables in similar situations. For example, one can approximate the uncertainty in the carcinogenic potency of an untested chemical by using the existing frequency distribution of potencies for chemicals already tested (Fiering et al., 1984).

Subjective Probability Distributions

A different method of probability assessment is based on expert opinion. In this method,

TABLE 9-2. Steps that could improve a quantitative uncertainty estimate.

1. *Determine the desired measure of risk* (e.g., mortality, life years lost, risk to the individual who is maximally exposed, number of persons at more than arbitrary "unacceptable" risk.) More than one measure will often be desired, but the remaining steps will need to be followed *de novo* for each method.
2. *Specify one or more "risk equations," mathematical relationships that express the risk measure in terms of its components.* For example, R = C x I x P (risk equals concentration times intake times potency) is a simple "risk equation" with three independent variables. Care must be taken to avoid both an excess and an insufficiency of detail.
3. *Generate an uncertainty distribution for each component.* This will generally involve the use of analogy, the use of statistical inference, of expert opinion, or a combination of these.
4. *Combine the individual distributions into a composite uncertainty distribution.* This step will often require Monte Carlo simulation (described later).
5. *"Recalibrate" the uncertainty distributions.* At this point, inferential analysis should enter or re-enter the process to corroborate or correct the outputs of step 4. In practice, it might involve altering the range of the distribution to account for dependence among the variables or truncating the distributions to exclude extreme values that are physically or logically impossible. Repeat steps 3, 4, and 5 as needed.
6. *Summarize the output, highlighting important implications for risk management.* Here the decision-maker and uncertainty analyst need to work together (or at least to understand each other's needs and limitations). In all written and oral presentations, the analyst should strive to ensure that the manager understands the following four aspects of the results:

- Their implications for supplanting any point estimate that might have been produced without consideration of uncertainty. In particular, presentations of uncertainty will help in advancing the debate over whether the standardized procedures used to generate point estimates of risk are too "conservative" in general or particular cases.
- Their insights regarding the balance between the costs of overestimating and underestimating risk (i.e., the shape and breadth of the uncertainty distribution informs the manager about how prudent various risk estimates might be).
- Their sensitivity to fundamentally unresolved scientific controversies.
- Their implications for research, identifying which uncertainties are most important and which uncertainties are amenable to reduction by directed research efforts. As part of this process, the analyst should attempt to quantify in absolute terms how much total effort might be put into reducing uncertainty before a control action is implemented (i.e., estimate the value of information using standard techniques).

Source: Adapted from Finkel, 1990.

TABLE 9-3. Some key variables in risk assessment for which probability distributions might be needed.

Model Component	Output Variable	Independent Parameter Variable
Transport	Air concentration	Chemical emission rate Stack exit temperature Stack exit velocity Mixing heights
Deposition	Deposition rate	Dry-deposition velocity Wet-deposition velocity Fraction of time with rain
Overland	Surface-water load	Fraction of chemical in overload runoff
Water	Surface-water concentration	River discharge Chemical decay coefficient in river
Soil	Surface-soil concentration	Surface-soil depth Exposure duration Exposure period Cation-exchange capacity Decay coefficient in soil
Food Chain	Plant concentration	Plant interception fraction Weathering elimination rate Crop density Soil-to-plant bioconcentration factor
	Fish concentration	Water-to-fish bioconcentration factor

the beliefs of selected experts are elicited and combined to provide a subjective probability distribution. This procedure can be used to estimate the uncertainty in a parameter (cf., the subjective assessment of the slope of the dose-response relationship for lead in Whitfield and Wallsten, 1989). However, subjective assessments are more often used for a risk assessment component for which the available inference options are logically or reasonably limited to a finite set of identifiable, plausible, and often mutually exclusive alternatives (i.e., for *model* uncertainty). In such an analysis, alternative scenarios or models are assigned subjective probability weights according to the best available data and scientific judgment; equal weights

might be used in the absence of reliable data or theoretical justifications supporting any option over any other. For example, this approach could be used to determine how much the risk assessor should rely on relative surface area vs. body weight in conducting a dose-response assessment. The application of particular sets of subjective probability weights in particular inference contexts could be standardized, codified, and updated as part of EPA's implementation of uncertainty analysis guidelines (see below).

Objective probabilities might seem inherently more accurate than subjective probabilities, but this is not always true. Formal methods (Bayesian statistics)[2] exist to incorporate objective information into a subjective probability distribution that reflects other matters that might be relevant but difficult to quantify, such as knowledge about chemical structure, expectations of the effects of concurrent exposure (synergy), or the scope of plausible variations in exposure. The chief advantage of an objective probability distribution is, of course, its objectivity; right or wrong, it is less likely to be susceptible to major and perhaps undetectable bias on the part of the analyst; this has palpable benefits in defending a risk assessment and the decisions that follow. A second advantage is that objective probability distributions are usually far easier to determine. However, there can be no rule that objective probability estimates are always preferred to subjective estimates, or vice versa.

Model Uncertainty: "Unconditional" Versus "Conditional" PDFs

Regardless of whether objective or subjective methods are used to assess them, the distinction between parameter uncertainty and model uncertainty remains pivotal and has implications for implementing improved risk assessments that acknowledge uncertainty. The most important difference between parameter uncertainty and model uncertainty, especially in the context of risk assessment, concerns how to interpret the output of an objective or subjective probability assessment for each.

One can readily construct a probability distribution for risk, exposure, potency, or some other quantity that reflects the probabilities that various values, corresponding to fundamentally different scientific models, represent the true state. Such a depiction, which we will call an "unconditional PDF" because it tries to represent all the uncertainty surrounding the quantity, can be useful for some decisions that agencies must make. In particular, EPA's research offices might be able to make more efficient decisions about where resources should be channeled to study particular risks, if the uncertainty in each risk were presented unconditionally. For example, an unconditional distribution might be reported in this way: "the potency of chemical X is 10^{-2} per part per million of air (with an uncertainty of a factor of 5 due to parameter uncertainty surrounding this value), but only if the LMS model is correct; if

[2] It is important to note that the distributions resulting from Bayesian models include various subjective judgments about models, data sets, etc. These are expressed as probability distributions but the probabilities should not be interpreted as probabilities of adverse effect but, rather, as expressions of strengths of conviction as to what models, data sets, etc. might be relevant to assessing risks of adverse effect. This is an important distinction which should be kept in mind when interpreting and using such distributions in risk management as a quantitative way of expressing uncertainty.

instead the chemical has a threshold, the potency at any ambient concentration is effectively zero." It might even help to assign subjective weights to the current thinking about the probability that each model is correct, especially if research decisions have to be made for many risks.

In addition, some specified *regulatory* decisions—those involving the ranking of different risks for the purpose of allowing "tradeoffs" or "offsets"—can also suffer if model uncertainty is not quantified. For example, two chemicals (Y and Z) with the same potency—assuming that the LMS model is correct—might involve different degrees of confidence in the veracity of that model assumption. If we judged that chemical Y had a 90%, or even a 20%, chance of acting in a threshold fashion, it might be a mistake to treat it as having the same potency as a chemical Z that is virtually certain to have no threshold and then to allow increased emissions of Z in exchange for greater reductions in Y.

However, unconditional statements of uncertainty can be misleading if managers use them for standard-setting, residual-risk decisions, or risk communication, and especially if others then misinterpret these statements. Consider two situations, involving the same hypothetical chemical, in which the same amount of uncertainty can have different implications, depending on whether it stems from parameter uncertainty (Situation A) or ignorance about model choice (Situation B). In Situation A, suppose that the uncertainty is due entirely to parameter sampling error in a single available bioassay involving few test animals. If 3 of 30 mice tested in that bioassay developed tumors, then a reasonable central-tendency estimate of the risk to mice at the dose used would be 0.1 (3/30). However, because of sampling error, there is approximately a 5% probability that the true number of tumors might be as low as zero (leading to zero as the lower confidence limit, LCL, of risk) and about a 5% probability that the true number of tumors is 6 or higher (leading to 0.2 (6/30) as the upper confidence limit, UCL, of risk).

In Situation B, suppose instead that the uncertainty is due entirely to ambiguity over which model of biological effect is correct. In this hypothetical situation, there was one bioassay in which 200 of 1,000 mice developed tumors; the risk to mice at that dose would be 0.2 (with essentially no parameter uncertainty due to the very large sample size). But suppose scientists disagree about whether the effect in mice is at all relevant to humans, because of profound metabolic or other differences between the two species, but can agree to assign equal probabilities of 50% to each eventuality. In this case as well, the LCL of the risk to humans would be zero (if the "nonrelevance" theory were correct), and the UCL would be 0.2 (if the "relevance" theory were correct), and it would be tempting to report a "central estimate" of 0.1, corresponding to the expected value of the two possible outcomes, weighted by their assigned probabilities. In either situation A or B, it would be *mathematically* correct to say the following: "The expected value of the estimate of the number of annual excess cancer deaths

nationwide caused by exposure to this substance is 1,000; the LCL of this estimate is zero deaths, and the UCL is 2,000 deaths.[3]"

We contend that in such cases, which typify the two kinds of uncertainties that risk managers must deal with, it would be a mistake simply to report the confidence limits and expected value in Situation B as one might do more routinely in Situation A, especially if one then used these summary statistics to make a regulatory decision. The risk-communication problem in treating this dichotomous model uncertainty (Situation B) as though it were a continuous probability distribution is that it obscures important information about the scientific controversy that must be resolved. Risk managers and the public should be given the opportunity to understand the sources of the controversy, to appreciate why the subjective weights assigned to each model are at their given values, and to judge for themselves what action is appropriate when the two theories, *at least one of which must be incorrect*, predict such disparate outcomes.

More critically, the expected value in Situation B might have dramatically different properties as an estimate for decision-making from the one in Situation A. The estimate of 1,000 deaths in Situation B is a contrivance of multiplying subjective weights that corresponds to no possible true value of risk, although this is not itself a fatal flaw; indeed, it is possible that a strategy of deliberately inviting errors of both overprotection and underprotection at each decision will turn out to be optimal over a long-run set of similar decisions. The more fundamental problem is that any estimate of central tendency does not necessarily lead to optimal decision-making. This would be true even if society had no desire to make conservative risk management decisions.

Simply put, although classical decision theory does encourage the use of expected values that take account of all sources of uncertainty, it is not in the decision-maker's or society's best interest *to treat fundamentally different predictions as quantities that can be "averaged" without considering the effects of each prediction on the decision that it leads to*. It is possible that a coin-toss gamble between zero deaths and 2,000 deaths would lead a regulator rationally to act as though 1,000 deaths were the certain outcome. But this is only a short-hand description of the actual process of expected-value decision-making, which asks how the *decisions* that correspond to estimates of zero deaths, 1,000 deaths, and 2,000 deaths perform relative to each other, in light of the possibility that each estimate (and hence each decision) is wrong. In other words, the choice to use an unconditional PDF when there is the kind of model uncertainty shown in situation B is a choice between the *possibility* of overprotecting or underprotecting—if one model is accepted and the other rejected—and the *certainty* of erring in one direction or the other if the hybrid estimate of 1,000 is constructed. Because in this example

---

[3]Assume that to convert from risk to the test animals to the predicted number of deaths in the human population, one must multiply by 10,000. Perhaps the laboratory dose is 10,000 times larger than the dose to humans, but 100 million humans are exposed. Thus, for example,

$$0.2 \left(\frac{risk}{laboratory\ dose} \right) \times 10^{-4} \left(\frac{laboratory\ dose}{ambient\ dose} \right) \times 10^{8} = 2000 \left(\frac{deaths}{ambient\ dose} \right).$$

the outcomes are numbers that can be manipulated mathematically, it is tempting to report the average, but this would surely be nonsensical if the outcomes were not numerical. If, for example, there were model uncertainty about where on the Gulf Coast a hurricane would hit, it would be sensible to elicit subjective judgment about the probability that a model predicting that the storm would hit in New Orleans was correct, versus the probability that an alternative model—say, one that predicted that the storm would hit in Tampa—was correct. It would also be sensible to assess the expected losses of lives and property if relief workers were irrevocably deployed in one location and the storm hit the other ("expected" losses in the sense of probability times magnitude). It would be foolish, however, to deploy workers irrevocably in Alabama on the grounds that it was the "expected value" of halfway between New Orleans and Tampa under the model uncertainty—and yet this is just the kind of reasoning invited by indiscriminate use of averages and percentiles from distributions dominated by model uncertainty.

Therefore, we recommend that analysts present *separate* assessments of the parameter uncertainty that remains for *each* independent choice of the underlying model(s) involved. This admonition is not inconsistent with our view that model uncertainty is important and that the ideal uncertainty analysis should consider and report all important uncertainties; we simply suspect that comprehension and decision-making might suffer if all uncertainties are lumped together indiscriminately. The subjective likelihood that each model (and hence each parameter uncertainty distribution) might be correct should still be elicited and reported, *but primarily to help the decision-maker gauge which depiction of risk and its associated parameter uncertainty is the correct one,* and not to construct a single hybrid distribution (except for particular purposes involving priority-setting, resource allocation, etc). In the hypothetical Situation B, this would mean presenting both models, their predictions, and their subjective weights, rather than simple summary statistics, such as the unconditional mean and UCL.

The existence of default options for model uncertainty (as discussed in the introduction to Part II and in Chapter 6) also places an important curb on the need for and use of unconditional depictions of uncertainty. If, as we recommend, EPA develops explicit principles for choosing and modifying its default models, it will further codify the practice that for every risk assessment, a sequence of "preferred" model choices will exist, with only one model being the prevailing choice at each inference point where scientific controversy exists. Therefore, the "default risk characterization," including uncertainty, will be the uncertainty distribution (embodying the various sources of parameter and scenario uncertainty) that is conditional on the approved choices for dose-response, exposure, uptake, and other models made under EPA's guidelines and principles. For each risk assessment, this PDF, rather than the single point estimate currently in force, should serve as the quantitative-risk input to standard-setting and residual-risk decisions that EPA will make under the act.

Thus, given the current state of the art and the realities of decision-making, model uncertainty should play only a subsidiary role in risk assessment and characterization, although it might be important when decision-makers integrate all the information necessary to make regulatory decisions. We recognize the intellectual and practical reasons for presenting alternative risk estimates and PDFs corresponding to alternative models that are scientifically

plausible, but that have not supplanted a default model chosen by EPA. However, we suggest that to create a single risk estimate or PDF out of various different models not only could undermine the entire notion of having default models that can be set aside for sufficient reason, but could lead to misleading and perhaps meaningless hybrid risk estimates. We have presented this discussion of the pitfalls of combining the results of incompatible models to support our view urging caution in applying these techniques in EPA's risk assessment. Such techniques should not be used for calculating unit risk estimates, because of the potential for misinterpretation of the quantitative risk characterization.[4] However, we encourage risk assessors and risk managers to work closely together to explore the implications of model uncertainty for risk management, and in this context explicit characterization of model uncertainty may be helpful. The characterization of model uncertainty may also be appropriate and useful for risk communication and for setting research priorities.

Finally, an uncertainty analysis that carefully keeps separate the influence of fundamental model uncertainties versus other types of uncertainty can reveal which controversies over model choice are actually important to risk management and which are "tempests in teapots." If, as might often be the case, the effect of all parameter uncertainties (and variabilities) is as large as or larger than that contributed by the controversy over model choice, then resolving the controversy over model choice would not be a high priority. In other words, if the "signal" to be discerned by a final answer as to which model or inference option is correct is not larger than the "noise" caused by parameter uncertainty in either (all) model(s), then effort should be focused on data collection to reduce the parameter uncertainties, rather than on basic research to resolve the modeling controversies.

SPECIFIC GUIDANCE ON UNCERTAINTY ANALYSIS

GENERATING PROBABILITY DISTRIBUTIONS

The following examples indicate how probability distributions might be developed in practice and illustrate many of the principles and recommended procedures discussed earlier in the chapter.

- *Example 1.* Estimated emission rates can differ significantly from actual values. Experience might show that emission estimates based on emission factors, mass balances, or material balances have an inherent uncertainty of a factor of about 100, whereas those based on testing tend to be within a factor of about 10. Expert opinion and analysis of past studies of

[4]Note that characterizing risks considering only the parameter uncertainty under the preferred set of models might not be as restrictive as it appears at first glance, in that some of the model choices can be safely recast as parameter uncertainties. For example, the choice of a scaling factor between rodents and humans need not be classified as a model choice between body weight and surface area that calls for two separate "conditional PDFs," but instead can be treated as an uncertain parameter in the equation $R_{human} \propto R_{rodent}\, BW^{a}$, where ^{a} might plausibly vary between 0.5 and 1.0 (see our discussion in Chapter 11). The only constraint in this case is that the scaling model is some power function of BW, the ratio of body weights.

such emission estimates could provide more definitive bounds on the estimates and result in a probability distribution. For example, a lognormal distribution with the median at the calculated emission estimate and a geometric standard deviation[5] of 10 (i.e., the case of emission factors) or $\sqrt{10}$ (for emissions based on testing).

• *Example 2.* A standard animal carcinogenicity bioassay provides the raw material for three related features of a complete uncertainty analysis. First, there is the random sampling uncertainty due to the limitation on the number of animals that can be tested. Suppose that at a particular dose 10 of 50 mice develop leukemia. The most likely estimate of the risk to each mouse would be calculated as 0.2 (the observed risk to the group, 10/50). However, chance dictates that if different groups of 50 animals were exposed to a risk of 0.2, some number n other than 10 might develop leukemia at each replication of the experiment. According to the binomial theorem, which governs independent dichotomous chance events (such as a coin falling either "heads" or "tails"), between 4 and 16 animals would develop cancer 99% of the time if many groups of 50 animals were exposed to identical lifetime risks of 0.2. EPA's standard procedure of reporting only the "q_1^*" value for potency is equivalent to computing the 95th percentile of random uncertainty using the binomial theorem (e.g., assuming that if 10 tumors were observed, 14 tumors would be a "conservative" estimate), and then finding the slope of the straight line drawn between this hypothetical response and the control-group response. Such a point estimate is informative neither about the plausible slopes greater and less than this value nor about the relative probabilities of the different plausible values. A distribution for q_1 derived from the entire binomial probability distribution for n, on the other hand, would answer both of these concerns.

A second opportunity, which allows the analyst to draw out some of the model uncertainty in dose-response relationships, stems from the flexibility of the LMS model. Even though this model is often viewed as unduly restrictive (e.g., it does not allow for thresholds or for "superlinear" dose-response relations at low doses), it is inherently flexible enough to account for sublinear dose-response relations (e.g., a quadratic function) at low doses. EPA's point-estimation procedure forces the q_1^* value to be associated with a linear low-dose model, but there is no reason why EPA could not fit an unrestricted model through all the values on the binomial uncertainty distribution of tumor response, thereby generating a distribution for potency that might include some probability that the true dose-response function is of quadratic or higher order (see, for example, Guess et al., 1977; Finkel, 1988).

Finally, EPA could account for another source of parameter uncertainty if it made use of more than one data set for each carcinogen. Techniques of meta-analysis, more and more frequently used to generate composite *point estimates* by averaging together the results of different studies (e.g., a second mouse study that might have found 20 leukemic animals out of 50 at the same dose), can perhaps more profitably be used to generate a composite *uncertainty*

---

[5]It is not always clear what percent of the distribution someone is referring to by "correct to within a factor of X." If instead of assuming that the person means with 100% confidence, we assumed that the person means 98% confidence, then the factor of X would cover two standard deviations on either side of the median, so one geometric standard deviation would be equal to $\sqrt{X}$.

distribution. This distribution could be broader than the binomial distribution that would arise from considering the sampling uncertainty in a single study, if the new study contradicted the first, or it could be narrower, if the results of each study were reinforcing (i.e., each result was well within the uncertainty range of the other).

• *Example 3.* The linearized multistage (LMS) model is often used to estimate dose-response relationships. Although many models could be used to estimate this relationship, two—the LMS and the biologically motivated (BM) models—seem to have the best biologic and mechanistic underpinning. Others, such as the probit and logit models, do not have a similar underpinning and are generic dose-response models. An additional possible advantage of BM models is their flexibility to accommodate the possibility of zero added response at low doses, even when there is a response at high doses. At present, there is rarely enough information to use BM models with great confidence, and a key issue is the plausibility of no increased hazard at low doses. If available information on such matters as biochemistry, genotoxicity, and induced cell replication suggests that low doses do not increase risk above background levels, then the question arises whether the subjective probability of risk at low doses should include both a positive probability that the risk is zero and a probability distribution for the degree of potency if it is not zero. In application, that might result in one of the following three decisions:

—If the data are sufficient to use the BM model, specify its parameters, and conclude scientifically (using whatever principles and evidentiary standards EPA sets forth in response to the committee's recommendation that it develop such principles) that this model is appropriate, the BM model could be used. Such occurrences are likely to be uncommon in the near term because of the need for extensive data of special types.

—If the data lead to a scientific conclusion that there is a substantial possibility that the low-dose potency is zero, the potency distributions from the BM and LMS models could be presented separately, perhaps with a narrative or quantitative statement of the probability weights to be assigned to each model.

—If the data do not suggest a substantial possibility of zero risk at low doses, the LMS model would continue to be used exclusively.

Statistical Analysis of Generated Probabilities

Once the needed subjective and objective probability distributions are estimated for each variable in the risk assessment, the estimates can be combined to determine their impact on the ultimate risk characterization. Joint distributions of input variables are often mathematically intractable, so an analyst must use approximating methods, such as numerical integration or Monte Carlo simulation. Such approximating methods can be made arbitrarily precise by appropriate computational methods. Numerical integration replaces the familiar operations of

integral calculus by summarizing the values of the dependent variable(s) on a very fine (multivariate) grid of the independent variables. Monte Carlo methods are similar, but sum the variables calculated at random points on the grid; this is especially advantageous when the number or complexity of the input variables is so large that the costs of evaluating all points on a sufficiently fine grid would be prohibitive. (For example, if each of three variables is examined at 100 points in all possible combination, the grid would require evaluation at 100^3 = 1,000,000 points, whereas a Monte Carlo simulation might provide results that are almost as accurate with only 1,000-10,000 randomly selected points.)

Barriers to Quantitative Uncertainty Analysis

The primary barriers to determining objective probabilities are lack of adequate scientific understanding and lack of needed data. Subjective probabilities are also not always available. For example, if the fundamental molecular-biologic bases of some hazards are not well understood, the associated scientific uncertainties cannot be reasonably characterized. In such a situation, it would be prudent public-health policy to adopt inference options from the conservative end of the spectrum of scientifically plausible available options. Quantitative dose-response assessment, with characterization of the uncertainty in the assessment, could then be conducted conditional on this set of inference options. Such a "conditional risk assessment" could then routinely be combined with an uncertainty analysis for exposure (which might not be subject to fundamental model uncertainty) to yield an estimate of risk and its associated uncertainty.

The committee recognizes the difficulties of using subjective probabilities in regulation. One is that someone would have to provide the probabilities to be used in a regulatory context. A "neutral" expert from within EPA or at a university or research center might not have the knowledge needed to provide a well-informed subjective probability distribution, whereas those who might have the most expertise might have or be perceived to have a conflict of interest, such as persons who work for the regulated source or for a public-interest group that has taken a stand on the matter. Allegations of conflict of interest or lack of knowledge regarding a chemical or issue might damage the credibility of the ultimate product of a subjective assessment. We note, however, that most of the same problems of real or perceived bias pervade EPA's current point-estimation approach.

At bottom, what matters is how risk managers and other end-users of risk assessments interpret the uncertainty in risk analysis. Correct interpretation is often difficult. For example, risks expressed on a logarithmic scale are commonly misinterpreted by assuming that an error of, say, a factor of 10 in one direction balances an error of a factor of 10 in the other. In fact, if a risk is expressed as 10^{-5} within a factor of 100 uncertainty in either direction, the average risk is approximately 1/2,000, rather than 1/100,000. In some senses, this is a problem of risk communication within the risk-assessment profession, rather than with the public.

UNCERTAINTY GUIDELINES

Contrary to EPA's statement that the quantitative techniques suggested in this chapter "require definition of the distribution of all input parameters and knowledge of the degree of dependence (e.g., covariance) among parameters," (EPA, 1991f) complete knowledge is not necessary for a Monte Carlo or similar approach to uncertainty analysis. In fact, such a statement is a tautology: it is the uncertainty analysis that tells scientists how their lack of "complete knowledge" affects the confidence they can have in their estimate. Although it is always better to be able to be precise about how uncertain one is, an imprecise statement of uncertainty reflects how uncertain the situation is—it is far better to acknowledge this than to respond to the "lack of complete knowledge" by holding fast to a "magic number" that one knows to be wildly overconfident. Uncertainty analysis simply estimates the logical implications of the assumed model and whatever assumed or empirical inputs the analyst chooses to use.

The difficulty in documenting uncertainty can be reduced by the use of uncertainty guidelines that will provide a structure for how to determine uncertainty for each parameter and for each plausible model. In some cases, objective probabilities are available for use. In others, a subjective consensus about the uncertainty may be based on whatever data are available. Once these decisions are documented, many of the difficulties in determining uncertainty can be alleviated. However, it is important to note that consensus might not be achieved. If a "first-cut" characterization of uncertainty in a specific case is deemed to be inappropriate or superseded by new information, it can be changed by means of such procedures as those outlined in Chapter 12.

The development of uncertainty guidelines is important, because a lack of clear statements as to how to address uncertainty in risk assessment might otherwise lead to continuing inconsistency in the extent to which uncertainty is explicitly considered in assessments done by EPA and other parties, as well as to inconsistencies in how uncertainty is quantified. Developing guidelines to promote consistency in efforts to understand the uncertainty in risk assessment should improve regulatory and public confidence in risk assessment, because guidelines would reduce inappropriate inconsistencies in approach, and where inconsistencies remain, they could help to explain why different federal or state agencies come to different conclusions when they analyze the same data.

RISK MANAGEMENT AND UNCERTAINTY ANALYSIS

The most important goal of uncertainty analysis is to improve risk management. Although the process of characterizing the uncertainty in a risk analysis is also subject to debate, it can at a minimum make clear to decision-makers and the public the ramifications of the risk analysis in the context of other public decisions. Uncertainty analysis also allows society to evaluate judgments made by experts when they disagree, an especially important attribute in a democratic society. Furthermore, because problems are not always resolved and analyses often

need to be repeated, identification and characterization of the uncertainties can make the repetition easier.

SINGLE ESTIMATES OF RISK

Once EPA succeeds in supplanting single point estimates with quantitative descriptions of uncertainty, its risk assessors will still need to summarize these distributions for risk managers (who will continue to use numerical estimates of risk as inputs to decision-making and risk communication). It is therefore crucial to understand that uncertainty analysis is *not* about replacing "risk numbers" with risk distributions or any other less transparent method; it is about *consciously* selecting the appropriate numerical estimate(s) from out of an understanding of the uncertainty.

Regardless of whether the applicable statute requires the manager to balance uncertain benefits and costs or to determine what level of risk is "acceptable," a bottom-line summary of the risk is a very important input, as it is critical to judging how confident the decision-maker can be that benefits exceed costs, that the residual risk is indeed "acceptable," or whatever other judgments must be made. Such summaries should include at least three types of information: (1) a fractile-based summary statistic, such as the median (the 50th percentile) or a 95th-percentile upper confidence limit, which denotes the probability that the uncertain quantity will fall an unspecified distance above or below some associated value; (2) an estimate of the mean and variance of the distribution, which along with the fractile-based statistic provides crucial information about how the probabilities and the absolute magnitudes of errors interrelate; and (3) a statement of the potential for errors and biases in these estimates of fractiles, mean, and variance, which can stem from ambiguity about the underlying models, approximations introduced to fit the distribution to a standard mathematical form, or both.

One important issue related to uncertainty is the extent to which a risk assessment that generates a point estimate, rather than a range of plausible values, is likely to be too "conservative" (that is, to excessively exaggerate the plausible magnitude of harm that might result from specified environmental exposures). As the two case studies that include uncertainty analysis (Appendixes F and G) illustrate, these investigations can show whether "conservatism" is in fact a problem, and if so, to what extent. Interestingly, the two studies reach opposite conclusions about "conservatism" in their specific risk-assessment situations; perhaps this suggests that facile conclusions about the "conservatism" of risk assessment in general might be off the mark. On the one hand, the study in Appendix G claims that EPA's estimate of MEI risk (approximately 10^{-1}) is in fact quite "conservative," given that the study calculates a "reasonable worst-case risk" to be only about 0.0015.[6] However, we note that this study essentially compared different and incompatible models for the cancer potency of butadiene, so it is impossible to discern what percentile of this unconditional uncertainty distribution any

[6]We arrive at this figure of 0.0015, or 1.5×10^{-3}, by noting that the "base case" for fenceline risk (Table 3-1 in Appendix G) is 5×10^{-4} and that "worst case estimates were two to three times higher than base case estimates."

estimate might be assigned (see the discussion of model uncertainty above). On the other hand, the Monte Carlo analysis of parameter uncertainty in exposure and potency in Appendix F claims that EPA's point estimate of risk from the coal-fired power plant was only at the 83rd percentile of the relevant uncertainty distribution. In other words, a standard "conservative" estimate of risk (the 95th percentile) *exceeds* EPA's value, in this case by a factor of 2.5. It also appears from Figure 5-7 in Appendix F that there is about a 1% chance that EPA's estimate is too low by more than a factor of 10. Note that both case studies (Appendixes F and G) fail to distinguish sources of uncertainty from sources of interindividual variability, so the corresponding "uncertainty" distributions obtained cannot be used to properly characterize uncertainty either in predicted incidence or in predicted risk to some particular (e.g., average, highly exposed, or high-risk) individual (see Chapter 11 and Appendix I-3).

As discussed above, access to the entire PDF allows the decision-maker to assess the amount of "conservatism" implicit in any estimate chosen from the distribution. In cases where the risk manager asks the analyst to summarize the PDF via one or more summary statistics, the committee suggests that EPA might consider a particular kind of point estimate to summarize uncertain risks, in light of the two distinct kinds of "conservatism" discussed in Appendix N-1 (the "level of conservatism," the relative percentile at which the point estimate of risk is located, and the "amount of conservatism," the absolute difference between the point estimate and the mean). Although the specific choice of this estimate should be left to EPA risk managers, and may also need to be flexible enough to accommodate case-specific circumstances, estimates do exist that can account for both the percentile and the relationship to the mean in one single number. For example, EPA could choose to summarize uncertain risks for reporting the mean of the upper five percent of the distribution. It is a mathematical truism that (for right-skewed distributions commonly encountered in risk assessment) the larger the uncertainty, the greater the chance that the mean may exceed any arbitrary percentile of the distribution (See Table 9-4). Thus, the mean of the upper five percent is by definition "conservative" both with respect to the overall mean of the distribution and to its 95th percentile, whereas the 95th percentile may not be a "conservative" estimate of the mean. In most situations, the amount of "conservatism" inherent in this new estimator will not be as extreme as it would be if a very high percentile (e.g. the 99.9th) was chosen without reference to the mean.

Thus, the issue of uncertainty subsumes the issue of conservatism in point estimates. Point estimates chosen without regard to uncertainty provide only the barest beginnings of the story in risk assessment. Excessive or insufficient conservatism can arise out of inattention to uncertainty, rather than out of a particular way of responding to uncertainty. Actions taken solely to reduce or eliminate potential conservatism will not reduce and might increase the problem of excessive reliance on point estimates.

In summary, EPA's position on the issue of uncertainty analysis (as represented in the Superfund document) seems plausible at first glance, but it might be somewhat muddled. If we know that "all risk numbers are only good to within a factor of 10," why do *any* analyses? The reason is that both the variance and the conservatism (if any) are case-specific and can

rarely be estimated with adequate precision until an honest attempt at uncertainty analysis is made.

Risk Communication

Inadequate scientific and technical communication about risk is sometimes a source of error and uncertainty, and guidance to risk assessors about what to include in a risk analysis should include guidance about how to present it. The risk assessor must strive to be understood (as well as to be accurate and complete), just as risk managers and other users must make themselves understood when they apply concepts that are sometimes difficult. This source of uncertainty in interprofessional communication seems to be almost untouched by EPA or any other official body (AIHC, 1992).

Comparison, Ranking, and Harmonization of Risk Assessments

As discussed in Chapter 6, EPA makes no attempt to apply a single set of methods to assess and compare default and alternative risk estimates with respect to parameter uncertainty. The same deficiency occurs in the comparison of risk estimates. When EPA ranks risks, it usually compares point estimates without considering the different uncertainties in each estimate. Even for less important regulatory decisions (when the financial and public-health impacts are deemed to be small), EPA should at least make sure that the point estimates of risk being compared are of the same type (e.g., that a 95% upper confidence bound for one risk is not compared with a median value for some other risk) and that each assessment has an informative (although perhaps sometimes brief) analysis of the uncertainty. For more important regulatory decisions, EPA should estimate the uncertainty in the *ratio* of the two risks and explicitly consider the probabilities and consequences of setting incorrect priorities. For any decisions involving risk-trading or priority-setting (e.g., for resource allocation or "offsets"), EPA should take into account information on the uncertainty in the quantities being ranked so as to ensure that such trades do not increase expected risk and that such priorities are directed at minimizing expected risk. When one or both risks are highly uncertain, EPA should also consider the probability and consequences of greatly erring in trading one risk for another, because in such cases one can lower the risk on average and yet introduce a small chance of greatly increasing risk.

Finally, EPA sometimes attempts to "harmonize" risk-assessment procedures between itself and other agencies, or among its own programs, by agreeing on a single common model assumption, even though the assumption chosen might have little more scientific plausibility than alternatives (e.g., replacing FDA's body-weight assumption and EPA's surface-area assumption with body weight to the 0.75 power). Such actions do not clarify or reduce the uncertainties in risk assessment. Rather than "harmonizing" risk assessments by picking one assumption over others when several assumptions are plausible and none of the assumptions is

clearly preferable, EPA should use the preferred models for risk calculation and characterization, but present the results of the alternative models (with their associated parameter uncertainties) to further inform decision-makers and the public. However, "harmonization" does serve an important purpose in the context of uncertainty analysis—it will help, rather than hinder, risk assessment if agencies cooperate to choose and validate a common set of *uncertainty distributions* (e.g., a standard PDF for the uncertain exponent in the "body weight to the X power" equation or a standard method for developing a PDF from a set of bioassay data).

FINDINGS AND RECOMMENDATIONS

The committee strongly supports the inclusion of uncertainty analysis in risk assessments despite the potential difficulties and costs involved. Even for lower-tier risk assessments, the inherent problems of uncertainty need to be made explicit through an analysis (although perhaps brief) of whatever data are available, perhaps with a statement about whether further uncertainty analysis is justified. The committee believes that a more explicit treatment of uncertainty is critical to the credibility of risk assessments and to their utility in risk management.

The committee's findings and recommendations are summarized briefly below.

SINGLE POINT ESTIMATES AND UNCERTAINTY

EPA often reports only a single point estimate of risk as a final output. In the past, EPA has only qualitatively acknowledged the uncertainty in its estimates, generally by referring to its risk estimates as "plausible upper bounds" with a plausible lower bound implied by the boilerplate statement that "the number could be as low as zero." In light of the inability to discern how "conservative" an estimate might be unless one does an uncertainty analysis, both statements might be misleading or untrue in particular cases.

- Use of a single point estimate suppresses information about sources of error that result from choices of model, data sets, and techniques for estimating values of parameters from data. EPA should not necessarily abandon the use of single-point estimates for decision-making, but such numbers must be the product of a consideration of both the estimate of risk and its uncertainties, not appear out of nowhere from a formulaic process. In other words, EPA should be free to choose a particular point estimate of risk to summarize the risk in light of its knowledge, uncertainty, and its desire to balance errors of overestimation and underestimation; but it should first derive that number from an uncertainty analysis of the risk estimate (e.g., using a summary statistic such as the "mean of the upper 5% of the distribution"). EPA should not simply state that its generic procedures yield the desired percentile. For example (although this is an analogous procedure to deal with variability, not uncertainty), EPA's

current way of calculating the "high-end exposure estimate" (see Chapter 10) is ad hoc, rather than systematic, and should be changed.

- EPA should make uncertainties explicit and present them as accurately and fully as is feasible and needed for risk management decision-making. To the greatest extent feasible, EPA should present quantitative, as opposed to qualitative, representations of uncertainty. However, EPA should not necessarily quantify model uncertainty (via subjective weights or any other technique), but should try to quantify the parameter and other uncertainty that exists for each plausible choice of scientific model. In this way, EPA can give its default models the primacy they are due under its guidelines, while presenting useful, but distinct alternative estimates of risk and uncertainty. In the quantitative portions of their risk characterizations (which will serve as one important input to standard-setting and residual-risk decisions under the Act), EPA risk assessors should consider only the uncertainty conditional on the choice of the preferred models for dose-response relationships, exposure, uptake, etc.
- In addition, uncertainty analyses should be refined only so far as improvements in the understanding of risk and the implications for risk management justify the expenditure of the professional time and other resources that are required.

Uncertainty Guidelines

EPA committed itself in a 1992 internal memorandum (see Appendix B) to doing some kind of uncertainty analysis in the future, but the memorandum does not define when or how such analysis might be done. In addition, it does not distinguish between the different types of uncertainty or provide specific examples. Thus, it provides only the first, critical step toward uncertainty analysis.

- EPA should develop uncertainty analysis guidelines—both a general set and specific language added to its existing guidelines for each step in risk assessment (e.g., the exposure assessment guidance). The guidelines should consider in some depth all the types of uncertainty (model, parameter, etc.) in all the stages of risk assessment. The uncertainty guidelines should require that the uncertainties in models, data sets, and parameters and their relative contributions to total uncertainty in a risk assessment be reported in a written risk-assessment document.

Comparison of Risk Estimates

EPA makes no attempt to apply a consistent method to assess and compare default and alternative risk estimates with respect to parameter uncertainty. Presentations of numerical values in an incomplete form lead to inappropriate and possibly misleading comparisons among risk estimates.

• When an alternative model is plausible enough to be considered for use in risk communication, or for potentially supplanting the default model when sufficient evidence becomes available, EPA should analyze parameter uncertainty at a similar level of detail for the default and alternative models. For example, in comparing risk estimates derived from delivered-dose versus PBPK models, EPA should qualify uncertainty in the interspecies scaling factor (for the former case) and in the parameters used to optimize the PBPK equations (for the latter case). Such comparisons may reveal that given current parameter uncertainties, the risk estimate chosen would not be particularly sensitive to the judgment about which model is correct.

Harmonization of Risk Assessment Methods

EPA sometimes attempts to "harmonize" risk-assessment procedures between itself and other agencies or among its own programs by agreeing on a single common model assumption, even though the assumption chosen might have little more scientific plausibility than alternatives, (e.g., replacing FDA's body-weight assumption and EPA's surface-area assumption with body weight to the 0.75 power). Such actions do not clarify or reduce the uncertainties in risk assessment.

• Rather than "harmonizing" risk assessments by picking one assumption over others when several assumptions are plausible and none of the assumptions is clearly preferable, EPA should maintain its own default assumption for regulatory decisions but indicate that any of the methods might be accurate and present the results as an uncertainty in the risk estimate or present multiple estimates and state the uncertainty in each. However, "harmonization" does serve an important purpose in the context of uncertainty analysis—it will help, rather than hinder, risk assessment if agencies cooperate to choose and validate a common set of *uncertainty distributions* (e.g., a standard *PDF* for the uncertain exponent in the "body weight to the X power" equation or a standard method for developing a PDF from a set of bioassay data).

Ranking of Risk

When EPA ranks risks, it usually compares point estimates without considering the different uncertainties in each estimate.

• For any decisions involving risk-trading or priority-setting (e.g., for resource allocation or "offsets"), EPA should take into account information on uncertainty in quantities being ranked so as to ensure that such trades do not increase expected risk and such priorities are directed at minimizing expected risk. When one or both risks are highly uncertain, EPA should also consider the probability and consequences of greatly erring in trading one risk for

another, because in such cases one can lower the risk on average and yet introduce a small chance of greatly increasing risk.

10
VARIABILITY

INTRODUCTION AND BACKGROUND

It is always difficult to identify the true level of risk in an endeavor like health risk assessment, which combines measurement, modeling, and inference or educated guesswork. Uncertainty analysis, the subject of Chapter 9, enables one to come to grips with how far away from the desired answer one's best estimate of an unknown quantity might be. Before we can complete an assessment of the uncertainty in an answer, however, we must recognize that many of our questions in risk assessment have *more than one useful answer*. Variability—typically, either across space, in time, or among individuals—complicates the search for the desired value of many important risk-assessment quantities.

Chapter 11 and Appendix I-3 discuss the issue of how to aggregate uncertainties and interindividual differences in each of the components of risk assessment. This chapter describes the sources of variability[1] and appropriate ways to characterize these interindividual differences in quantities related to predicted risk.

Variability is a very well-known "fact of life" in many fields of science, but its sources, effects, and ramifications are not yet routinely appreciated in environmental health risk assessment and management. Accordingly, the first section of this chapter will step back and deal with the general phenomenon (using some examples relevant to risk assessment, but not exclusively), and then for the remainder of the chapter focus only on variability in quantities that directly influence calculations of individual and population risk.

When an important quantity is both uncertain and variable, opportunities are created to fundamentally misunderstand or misestimate the behavior of the quantity.

To draw an analogy, the exact distance between the earth and the moon is both difficult to measure precisely (at least it was until the very recent past) and changeable, because the moon's orbit is elliptical, rather than circular. Thus, as seen in Figure 10-1, uncertainty and variability can complement or confound each other. When only scattered measurements of the

[1]Some specialists in different fields often use the term "variability" to refer to a dispersion of possible or actual values associated with a particular quantity, often with reference to random variability associated with any estimate of an unknown (i.e., uncertain) quantity. This report, unless stated otherwise, will use the terms interindividual variability, variability, and interindividual heterogeneity all to refer to individual-to-individual differences in quantities associated with predicted risk, such as in measures of or parameters used to model ambient concentration, uptake or exposure per unit ambient concentration, biologically effective dose per unit exposure, and increased risk per unit effective dose.

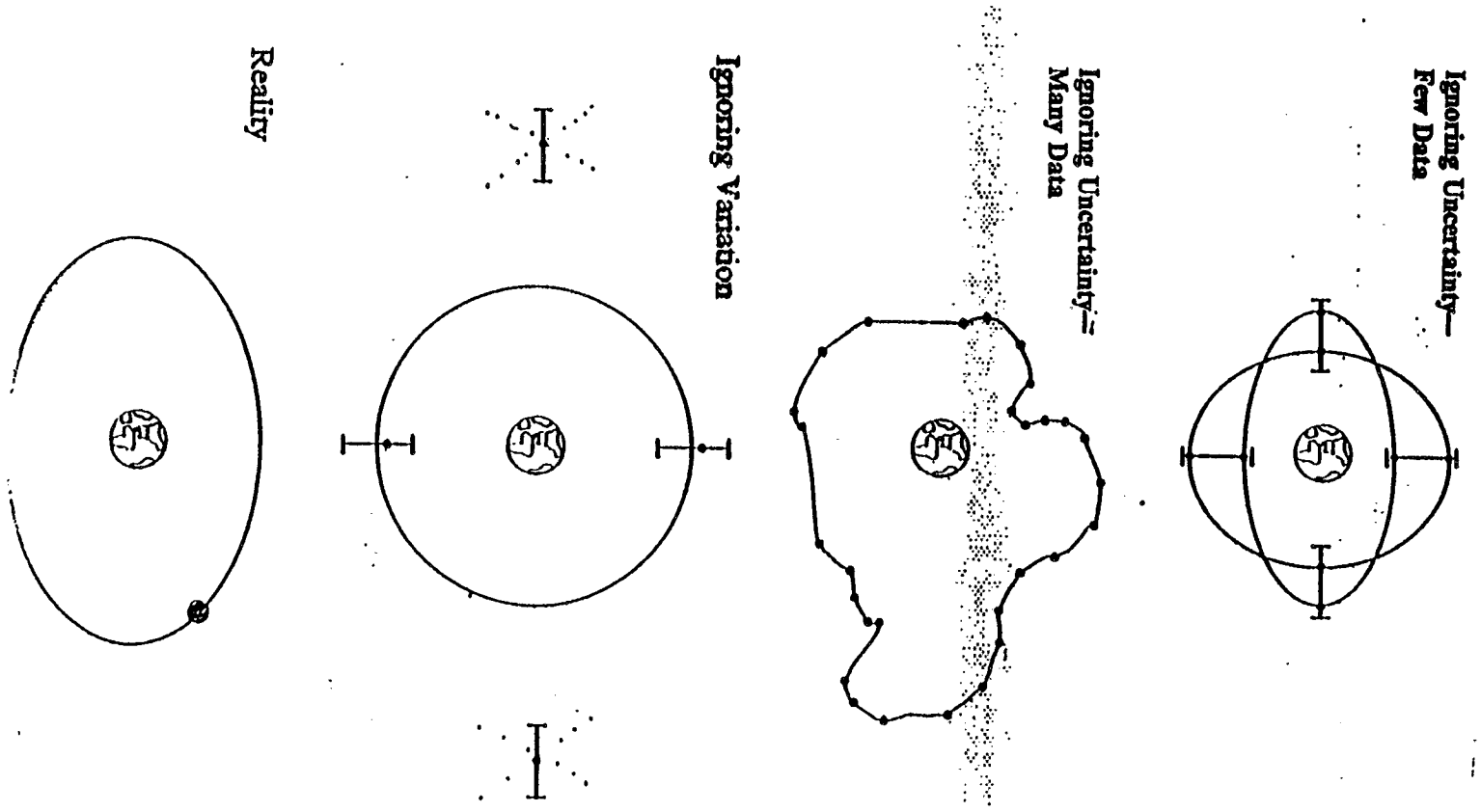


Figure 10-1. Effects of ignoring uncertainty versus ignoring variability in measuring the distance between the earth and the moon.

earth-moon distance were available, the variation among them might have led astronomers to conclude that their measurements were faulty (i.e., ascribing to uncertainty what was actually caused by variability) or that the moon's orbit was random (i.e., not allowing for uncertainty to shed light on seemingly unexplainable differences that are in fact variable *and* predictable). The most basic flaw of all would be to simply misestimate the true distance (the third diagram in Figure 10-1) by assuming that a few observations were sufficient (after correcting for measurement error, if applicable). This is probably the pitfall that is most relevant for health risk assessment: treating a highly variable quantity as if it was invariant or only uncertain, thereby yielding an estimate that is incorrect for some of the population (or some of the time, or over some locations), or even one that is also an inaccurate estimate of the average over the entire population.

In the risk-assessment paradigm, there are many sources of variability. Certainly, the regulation of air pollutants has long recognized that chemicals differ from each other in their physical and toxic properties and that sources differ from each other in their emission rates and characteristics; such variability is built into virtually any sensible question of risk assessment or control. However, even if we focus on a single substance emanating from a single stationary source, variability pervades each stage from emission to health or ecologic end point:

- *Emissions* vary temporally, both in flux and in release characteristics, such as temperature and pressure.
- The *transport and fate* of the pollutant vary with such well-understood factors as wind speed, wind direction, and exposure to sunlight (and such less-acknowledged factors as humidity and terrain), so its concentrations around its source vary spatially and temporally.
- Individual human *exposures* vary according to individual differences in breathing rates, food consumption, and activity (e.g., time spent in each microenvironment).
- The *dose-response* relationship (the "potency") varies for a single pollutant, because each human is uniquely susceptible to carcinogenic or other stimuli (and this inherent susceptibility might well vary during the lifetime of each person, or vary with such things as other illness or exposures to other agents).

Each of these variabilities is in turn often composed of several underlying variable phenomena. For example, the natural variability in human weight is due to the interaction of genetic, nutritional, and other environmental factors. According to the central limit theorem, variability that arises from independent factors that act multiplicatively will generally lead to an approximately lognormal distribution across the population or spatial/temporal dimension (as is commonly observed when concentrations of air pollutants are plotted).

When there is more than one desired answer to a scientific question where the search for truth is the end in itself, only two responses are ultimately satisfactory: gather more data or rephrase the question. For example, the question "How far away is the moon from the earth?" cannot be answered both simply and correctly. Either enough data must be obtained to give an answer of the form "The distance ranges between 221,460 and 252,710 miles" or "The moon's orbit is approximately elliptical, with a minor axis of 442,920 miles, a major axis of 505,420

miles, and an eccentricity of 0.482," or the question must be reduced to one with a single right answer (e.g., "How far away is the moon from the earth *at its perigee?*").

When the question is not purely scientific, but is intended to support a social decision, the decision-maker has a few more options, although each course of action will have repercussions that might foreclose other courses. Briefly, variability in the substance of a regulatory or science-policy question can be dealt with in four basic ways:

1. *Ignore the variability and hope for the best.* This strategy tends to be most successful when the variability is small and any estimate that ignores it will not be far from the truth. For example, the Environmental Protection Agency's (EPA's) practice of assuming that all adults weigh 70 kg is likely to be correct to within ±25% for most adults and probably valid to within a factor of 3 for virtually all adults. However, this approach may not be appropriate for children, where variability may be large (NRC, 1993e).

2. *Explicitly disaggregate the variability.* Where the quantity seems to change smoothly and predictably over some range, continuous mathematical models may be fitted to the data in place of a discrete step function. An example might be the fitting of sine waves to annual concentration cycles for a particular pollutant. In other cases, it is easier to disaggregate the data by considering all or the relevant subgroups or subpopulations. For interindividual variability, this involves dividing the population into as many subpopulations as deemed necessary. For example, one might perform a separate risk assessment for short-term exposure to high levels of ionizing radiation for each 10-year age interval in the population, to take account of age-related differences in susceptibility. For temporal variability, it involves modeling or measuring in a discrete, rather than a continuous, fashion, on an appropriate time scale. For example, a specific type of air-pollution monitor might collect air for 15 min of each hour and report the 15-min average concentration of some pollutant. Such values might then be further aggregated to produce summary values at an even coarser time scale. For spatial variability, it involves choosing an appropriate subregion, e.g., modeling the extent of global warming or cooling for each 10-deg swath of latitude around the globe, rather than predicting a single value for the entire planet, which might mask substantial and important regional differences. In each case, the common thread appears: when variability is "large" over the entire data set, the variability within each subset can become sufficiently "small" ("small" in the sense of the body-weight example in the paragraph above), if the data are disaggregated into an appropriate number of qualitatively distinct subsets. The strategy tends to be most successful when the stakes are so high (or the data or estimates so easy to obtain) that the proliferation of separate assessments does not consume inordinate amounts of resources. In contrast, in studies of a phenomenon such as global climate change, where the stakes are quite high, the estimates may also be quite hard to obtain on a highly disaggregated basis.

In health risk assessment, the choice of the averaging time used to transform the variable quantity into a more manageable form is crucially important. In general, for the assessment of acute toxicity, estimates of the variability in exposure and/or uptake over relatively short periods (minutes, hours, days) are needed. For chronic effects such as cancer, one might

model exposure and/or update over months or years without losing needed information, since short-term "peaks and valleys" would matter for cancer risk assessment only insofar as they affected the long-term or lifetime average exposure.[2] The longer-term variability will generally, though not always, be significantly less marked than the variation over the short-term (but see footnote 3). Moreover, the shorter the averaging time, the more such periods will be contained in an individual's lifetime, and the more opportunity there will be for rare fluctuations in exposure or uptake to produce significant risks. This, for example, explains why regulators concerned with the health effects of tropospheric ozone consider the combination of peak short-term concentration and peak activity (e.g., the "exercising asthmatic"). In all cases, the exposure assessor needs to determine which time periods are relevant for which toxic effects, and then see whether available data measuring exposure, uptake, internal dose rates, etc., can provide estimates of both the average and the variability over the necessary averaging time.

3. *Use the average value of a quantity that varies.* This strategy is not the same as ignoring the variability; ideally, it follows from a decision that the average value can be *estimated reliably in light of the variability,* and that it is a good surrogate for the variable quantity. For example, EPA often uses 70 kg as the average body weight of an adult, presumably because although many adults weigh as little as 40 kg and as much as 100 kg, the average weight is almost as useful as (and less complicated than) three different "scenario" values or an entire distribution of weights. In the same vein, a layperson might be content in knowing the average value of the moon's distance from the earth, rather than the minimum, average, and maximum (let alone a complete mathematical description of its orbit)—whereas the average alone would be useless, or even dangerous, to the National Aeronautics and Space Administration in planning an Apollo mission. Thus, this strategy tends to be most successful (and indeed might be the only sensible strategy) when the variability is small[3] or when the quantity is itself an input for a model or decision in which the average value of the end result (the combination of several quantities) is all that matters, either for scientific or policy reasons. An example of a scientific rationale for using the average value is the long-term average

---

[2]This assumes that risk is linear in long-term average dose, which is one of the bases of the classical models of carcinogenesis (e.g., the LMS dose-response model using administered dose). However, when one moves to more sophisticated models of the dose-exposure (i.e., PBPK) and exposure-response (i.e., biologically motivated or cell-kinetics models) relationships, shorter averaging times become important even though the health endpoint may manifest itself over the long-term. For example, the cancer risk from a chemical that is both metabolically activated and detoxified in vivo may not be a function of total exposure, but only of those periods of exposure during which detoxification pathways cannot keep pace with activating ones. In such cases, data on average long-term concentrations (and interindividual variability therein) may completely miss the only toxicologically relevant exposure periods.

[3]As discussed above, in many cases variability that exists over a short averaging time may grow less and less important as the averaging time increases. For example, if on average, adults breathe 20m^3 of air per day, then over any random 1-minute period, in a group of 1,000 adults there would probably be some (those involved in heavy exertion) breathing much more than the average value of 0.014 (m^3/min), and other (those asleep) breathing much less. Over the course of a year, however, the variation around the average value of 7300 m^3/yr would be much smaller, as periods of heavy exercise, sleep, and average activity "average out." On the other hand, some varying human characteristics do not substantially converge over longer averaging periods. For example, the daily variation in the amount of apple juice people drink probably mirrors the monthly and yearly variation as well—those individuals who drink no apple juice on a random day are probably those who rarely or never drink it, while those at the other "tail" of the distribution (drinking perhaps three glasses per day) probably tend to repeat this pattern day after day (in other words, the distribution of "glasses drunk per year" probably extends all the way from zero to 365x3, rather than varying narrowly around the midpoint of this range).

concentration of a carcinogen in air. If the dose-response function is linear (i.e., "potency" is a single number), the end result (risk) is proportional to the average concentration. If the concentration is, say, 10 ppm higher than the average in one week and 10 ppm lower than the average in another week, this variability will have no effect on an exposed person's lifetime risk, so it is biologically unimportant. An example of a policy rationale is the use of the expected number of cancer cases in a population exposed to varying concentrations of an airborne carcinogen. If it is determined for a particular policy rationale that the distribution of individual risks across the population does not matter, then the product of *average* concentration, potency and population size equals the expected incidence, and the spread of concentrations about the average concentration is similarly unimportant. The average value is also the summary statistic of choice for social decisions when there is an opportunity for errors of underestimation and overestimation (which lead to underregulation and overregulation) to even out over a large set of similar choices over the long run.

There are at least two reasons why large variabilities can lead to precarious decisions if the average value is used. The obvious problem is that individual characteristics of persons or situations far from the average are "averaged away" and can no longer be identified or reported. A less obvious pitfall occurs when the variability is dichotomous (or has several discrete values) and the average corresponds to a value that does not exist in nature. If men and women respond markedly differently to some exposure situation, for example, the decision that would be appropriate if there existed an "average person" (midway between man and woman) might be inappropriate for either category of real person (see Finkel, 1991).

4. *Use a maximum or minimum of a quantity that varies.* This is perhaps the most common way of dealing with variability in risk assessment—to focus attention on one period (e.g., the period of peak exposure), one spatial subregion (e.g., the location where the "maximally exposed individual" resides), or one subpopulation (e.g., exercising asthmatics or children who ingest pathologically large amounts of soil) and ignore the rest. This strategy tends to be most successful when the measures needed to protect or account for the person (or situation) with the extreme value will *also* suffice for the remainder of the distribution. It is also important to ensure that this strategy will not impose inordinate costs, compared with other approaches (such as using different controls for each subregion or population or simply controlling less stringently by using the average value instead of the extreme "tail").

The crucial point to bear in mind about all four of those strategies for dealing with variability is that unless someone measures, estimates, or at least roughly models the extent and nature of the variability, any strategy will be precarious. It stands to reason that strategy 1 ("hope for the best") hinges on the assumption that the variability is small—an assumption whose verification requires at least some attention to variability. Similarly, strategy 2 requires the definition of subregions or subpopulations in each of which the variability is small, so care must be taken to avoid the same conundrum that applies to strategy 1. (It is difficult to be sure that you can ignore variability until you think about the possible consequences of ignoring it). Less obviously, one still needs to be somewhat confident that one has a handle on the variability in order to reduce the distribution to either an average (strategy 3) or a "tail" value

(strategy 4). We *know* that 70 kg is an average adult body weight (and that virtually no adults are above or below 70 kg by more than a factor of 3), because weight is directly observable *and* because we know the mechanism by which people grow and the biologic limits of either extreme. Armed with our senses and this knowledge, we might need only a few observations to pin down roughly the minimum, the average, and the maximum. But what about a variable like "the rate at which human liver cells metabolize ethylene dibromide into its glutathione conjugate"? *Here a few direct measurements or a few extrapolations from animals may not be adequate,* because in the absence of any firm notion of the spread of this distribution within the human population (or the mechanisms by which the spread occurs), we cannot know how reliably our estimate of the average value reflects the true average, nor how well the observed minimum and maximum mirror the *true* extremes.

The distribution for an important variable such as metabolic rate should thus explicitly be considered in the risk assessment, and the reliability of the overall risk estimate should reflect knowledge about both the uncertainty and the variability in this characteristic. The importance of a more accurate risk estimate may motivate additional measurements of this variable, so that its distributions may be better defined with these additional data.

This chapter concentrates on how EPA treats variability in emissions, exposures, and dose-response relationships, to identify which of the four strategies it typically uses and to assess how adequately it has considered each choice and its consequences. The goals of this chapter are three: (1) to indicate how EPA can increase its sophistication in defining variability and handling its effects; (2) to provide information as to how to improve risk *communication,* so that Congress and the public understand at least which variabilities are and which are not accounted for, and how EPA's handling of variability affects the "conservatism" (or lack thereof) inherent in its risk numbers; and (3) to recommend specific research whose results could lead to useful changes in risk-assessment procedures.

In recent years, EPA has begun to increase its attention to variability. Moreover, the lack of attention in the past was due in part to a set of choices to erect a set of conservative default options (strategy 4 above) instead of dealing with variability explicitly. In theory at least, the question "How do you determine the extreme of a distribution without knowing the whole distribution?" can be answered by setting a highly conservative default and placing the burden of proof on those who wish to relax the default by showing that the extreme is unrealistic even as a "worst case." For example, the concept of the MEI (someone who breathes pollutants from the source for 70 years, 24 hours per day, at a specified location near a plant boundary) has been criticized as unrealistic, but most agree that as a summary of the population distribution of "number of hours spent at a given location during a lifetime" it might be a reasonable place to start from as a conservative short-cut for the entire distribution.

EPA has also tackled interindividual variability squarely in *Exposure Factors Handbook* (EPA, 1989c), which provides various percentiles (e.g., 5th, 25th, 50th, 75th, 95th) of the observed variability distributions for some components of exposure assessment, such as breathing rates, water ingestion, and consumption of particular foodstuffs. This document has not yet become a standard reference for many of EPA's offices, however. In addition, as we will discuss below, EPA has not dealt adequately with several other major sources of variabil-

ity. *As a result, EPA's methods to manage variability in risk assessment rely on an ill-characterized mix of some questionable distributions, some verified and unverified point values intended to be "averages," some verified and unverified point values intended to be "worst cases," and some "missing defaults," that is, hidden assumptions that ignore important sources of variability.*

Moreover, several trends in risk assessment and risk management are now increasing the urgency of a broad and well-considered strategy to deal with variability. The three most important of these trends are the following:

- *The emergence of more sophisticated biological models for risk assessment.* As pharmacokinetic models replace the administered assumption and as cell-kinetics models (such as the Moolgavkar-Venzon-Knudson model) replace the linearized-multistage model, default models that ignored human variability or took conservative measures to sidestep it will be supplanted by models that explicitly contain values of biologic measures intended to represent the human population. If the latter models ignore variability or use unverified surrogates for presumed average or worst-case properties, risk assessment might take a step backwards, becoming either less or more conservative without anyone's knowledge.
- *The growing interest in detailed assessments of the actual exposures that people face, rather than hypothetical worst-case exposures.* To be trustworthy, both average and worst-case surrogates for variability require some knowledge of the rest of the distribution, as mentioned above. However, it is not well recognized that the average might be *more* sensitive to the extreme portions of the whole distribution than an upper percentile might be, such as the 95th. In addition, the use of such terms as *actual* and *best estimates* carries an expectation of precision that might apply to only *part* of the exposure assessment, dose-response relationship, or risk assessment. If, for example, we could precisely measure the airborne concentration of a pollutant in a community around a stationary source (i.e., understand the spatial variability), but did not know the population distribution of breathing rates, we could not predict *anyone's* "actual exposure." In fact, even if we knew *both* distributions but could not superimpose them (i.e., know which breathing rates went with which concentrations), the predictions would be as variable as either of the underlying distributions. These circumstances speak to the need for progress in many kinds of research and data collection at once, if we wish to improve the power and the realism of risk assessment.
- *The growing interest in risk-reduction measures that target people, rather than sources.* It should go without saying that if government or industry wishes to eliminate unacceptably high risks to particular persons by purchasing their homes, providing them with bottled water, restricting access to "hot spots" of risk, etc., it needs to know precisely who those persons are and where or when those hot spots are occurring. Even if such policies were not highly controversial and difficult to implement in an equitable and socially responsive way, merely identifying the prospective targets of such policies may well presuppose a command of variability beyond our current capabilities.

EXPOSURE VARIABILITY

Variability in human response to pollutants emitted from a particular source or set of sources can arise from differences in characteristics of exposure, uptake, and personal dose-response relationships (susceptibility). Exposure variability in turn depends on variability in all the factors that affect exposure, including emissions, atmospheric processes (transport and transformation), personal activity, and the pollutant concentration in the microenvironments where the exposures occur. Information on those variabilities is not routinely included in EPA's exposure assessments, probably because it has been difficult to specify the distributions that describe the variations.

Human exposure results from the contact of a person with a substance at some nonzero concentration. Thus, it is tied to personal activities that determine a person's location (e.g., outdoors vs. indoors, standing downwind of an industrial facility vs. riding in a car, in the kitchen vs. on a porch); the person's level of activity and breathing rate influences the uptake of airborne pollutants. Exposure is also tied to emission rates and atmospheric processes that affect pollutant concentrations in the microenvironment where the person is exposed. Such processes include infiltration of outside air indoors, atmospheric advection (i.e., transport by the prevailing wind), diffusion (i.e., transport by atmospheric turbulence), chemical and physical transformation, deposition, and re-entrainment—variability in each process tends to increase the overall variability in exposure. The variabilities in emissions atmospheric processes, characteristics of the microenvironment, and personal activity are not necessarily independent of each other; for example, personal activities and pollutant concentrations at a specific location might change in response to outdoor temperature; they might also differ between weekends and weekdays because the level of industrial activity changes.

EMISSIONS VARIABILITY

There are basically four categories of emission variability that may need separate assessment methods, depending on the circumstances:

- Routine—this is the type most frequently covered by current approaches.
- Ordinary maintenance—special emissions may occur, for example, when the bag house is cleaned. In other cases certain emissions may only occur during maintenance, as when a specific volatile cleaner is routinely used to scour or wash out a reaction tank. These can be deliberately observed and monitored to obtain needed emissions information, if this mode is deemed likely to be significant.
- Upsets and breakdowns—unusual operating conditions that may recur within average periods of days, weeks, or months, depending on the facility/process. A combination of observations and modeling approaches may be needed here.
- Catastrophic failures—large explosions, ruptures of storage tanks, etc.

The last category is addressed in a separate section of the Clean Air Act and is not discussed in this report.

At least two major factors influence variability in emissions as it affects exposure assessment. First, a given source typically does not emit at a constant rate. It is subject to such things as load changes, upsets, fuel changes, process modifications, and environmental influences. Some sources are, by their nature, intermittent or cyclical. A second factor is that two similar sources (e.g., facilities in the same source category) can emit at different rates because of differences in such things as age, maintenance or production details.

The automobile is an excellent example of both causes. Consider a single, well-characterized car with an effective control system. When it is started, the catalyst has not warmed up, and emissions can be high. Almost half the total automobile emissions in, say, Los Angeles can occur during the cold-start period. After the catalyst reaches its appropriate temperature range, it is extremely effective (>90%) at removing organic substances, such as benzene and formaldehyde, during most of the driving period. However, hard accelerations can overwhelm the system's capabilities and lead to high emissions. Those variations can lead to spatial and temporal distributions of emissions in a city (e.g., high emissions in areas with a large number of cold starts, particularly in the morning). The composition of the emissions, including the toxic content, differs between cold-start and driving periods. Emissions also differ between cars—often dramatically. Because of differences in control equipment, total emissions can vary, and emissions between cycles can vary between cars (e.g., cold-start vs. evaporative emissions). A final notable contribution to emission variability in automobiles is the presence of super-emitters, whose control systems have failed and may emit organic substances at a rate 10 times that of a comparable vehicle that is operating properly.

Thus, an exposure analysis based on source-category average emissions will miss the variability in sources within that category. And, exposure analyses that do not account for temporal changes in emissions from a particular source will miss an important factor, especially to the extent that emissions are linked to meteorologic conditions. In many cases, it is difficult or impossible to know a priori how emissions will vary, particularly because of upsets in processes that could lead to high exposures over short periods.

Atmospheric Process Variability

Meteorologic conditions greatly influence the dispersion, transformation, and deposition of pollutants. For example, ozone concentrations are highest during summer afternoons, whereas carbon monoxide and benzene concentrations peak in the morning (because of the combination of large emissions and little dilution) and during the winter. Formaldehyde can peak in the afternoon during the summer (because of photochemical production) and in the morning in the winter (because of rush-hour emissions and little dilution). Concentrations of primary (i.e., emitted) pollutants, such as benzene and carbon monoxide, are higher in the winter in urban areas, whereas those of many secondary pollutants (i.e., those resulting from atmospheric transformations of primary pollutants), such as ozone, are higher in the summer. Meteoro-

logic conditions may also play a role in regional variations. Some areas experience long periods of stagnant air, which lead to very high concentrations of both primary and secondary pollutants. An extreme example is the London smog that led to high death rates before the mid-1950s. Wind velocity and mixing height also influence pollutant concentrations. (Mixing height is the height to which pollutants are rapidly mixed due to atmospheric turbulence; in effect, it is one dimension of the atmospheric volume in which pollutants are diluted.) They are usually correlated; the prevailing winds and velocities in the winter, when the mixing height is low, can be very different from those in the summer.

Some quantitative information is available about the impact of meteorologic variability on pollutant concentrations. Concentrations measured at one location over some period tend to follow a lognormal distribution. There are significant fluctuations in the concentrations about the medians (e.g., Seinfeld, 1986), which often vary by a factor of more than 10. The extreme concentrations are usually related to time and season. The relative magnitudes and frequencies of such fluctuations in concentration increase as distance from the source decreases. Pollutant transport over complex terrain (e.g., presence of hills or tall buildings), which is generally difficult to model, can further increase relative differences in extreme concentrations about the medians. Two examples of the influence of complex terrain are Donora, Pennsylvania (in a river valley), and the Meuse Valley in Belgium. In those areas, as in London, periods of extremely high pollutant concentrations led to a period of increased deaths. Estimates of concentration over flat terrain cannot capture such effects.

Empirical data on concentration variability are sparse, except for a few pollutants, notably the criteria pollutants (including carbon monoxide, ozone, sulfur dioxide, and particulate matter). Some information on variations in formaldehyde and benzene concentrations is also available. One interesting study that considered air-pollutant exposure during commuting in the Los Angeles area was conducted by the South Coast Air Quality Management District (SCAQMD, 1989). The authors looked at exposure dependence on seasonal, vehicular-age, and freeway-use variations. They found that drivers of older vehicles had greater exposure to benzene and that exposure to benzene, formaldehyde, ethylene, and chromium was greater in the winter, although exposure to ethylene dichloride was greater in the summer. They did not report the variability in exposure between similar vehicles or distributions of the exposures (e.g., probability density functions).

Microenvironmental and Personal-Activity Variability

Microenvironmental variability, particularly when compounded with differences in personal activity, can contribute to substantial variability in individual exposure. For example, the lifetime-exposed 70-year-old has been faulted as an extreme case, but it is instructive to consider this hypothetical person in the distribution of personal activity traits. Although it is unlikely, this 70-year lifetime exposure activity pattern is one end of the spectrum in the variability of personal activity and time spent in a specific microenvironment.

Concentrations in various microenvironments vary considerably and depend on a variety of

factors, such as species, building type, ventilation system, locality of other sources, and street canyon width and depth. Both the Los Angeles study (SCAQMD, 1989) and a New Jersey study (Weisel et al., 1992) revealed that exposure can be increased during commuting, particularly if the automobile itself is defective. The primary sources of many air pollutants are indoors, so their highest concentrations are found there. Those concentrations can be 10-1,000 times the outdoor concentrations (or even greater). However, the difference between outdoor and indoor concentrations of pollutants is not nearly so great when the indoor location is ventilated. Concentrations of compounds that do not react rapidly with or settle on surfaces, such as carbon monoxide and many organic compounds might not decrease significantly when ventilated indoors. If there are additional sources of these compounds indoors, their concentrations might, in fact, increase. Concentrations of more reactive compounds, such as ozone, can decrease by a factor of 2 or more, depending on ventilation rate and the ventilation system used (Nazaroff and Cass, 1986). Particles can also be advected indoors (Nazaroff et al., 1990). One concern is that the ventilation of outdoor pollutants indoors can increase the formation of other pollutants (Nazaroff and Cass, 1986; Weschler et al., 1992). The lifetime-exposed person sitting on the porch outside his home may be at one extreme for exposure to emissions from an outdoor stationary source, but may be at the other extreme for net air-pollutant exposure; such a person may have effectively avoided "hot" microenvironments in both the home and the automobile.

Increased personal activity leads to a larger uptake, and this will add to variability by as much as a factor of about 2 or more. The activity-related component of variability depends on both the microenvironmental variability (e.g., outdoors vs. indoors) and personal characteristics (e.g., children vs. adults).

VARIABILITY IN HUMAN SUSCEPTIBILITY

Person-to-person differences in behavior, genetic makeup, and life history together confer on individual people unique susceptibilities to carcinogenesis (Harris, 1991). Such inter-individual differences can be inherited or acquired. For example, inherited differences in susceptibility to physical or chemical carcinogens have been observed, including a substantially increased risk of sunlight-induced skin cancer in people with xeroderma pigmentosum, of bladder cancer in dyestuff workers whose genetic makeup results in the "poor acetylator" phenotype, and of bronchogenic carcinoma in tobacco smokers who have an "extensive debrisoquine hydroxylator" phenotype (both are described further in Appendix H). Similarly among different inbred and outbred strains of laboratory animals (and within particular outbred strains) exposed to carcinogenic initiators or tumor promoters there may be a factor of 40 variation in tumor response (Boutwell, 1964; Drinkwater and Bennett, 1991; Walker et al., 1992). Acquired differences that can significantly affect an individual's susceptibility to carcinogenesis include the presence of concurrent viral or other infectious diseases, nutritional factors such as alcohol and fiber intake, and temporal factors such as stress and aging.

Appendix H describes three classes of factors that can affect susceptibility: (1) those which

are rare in the human population but which confer very large increases in susceptibility upon those affected; (2) those which are very common but only marginally increase susceptibility; and (3) those which may be neither rare nor of marginal importance to those affected. The Appendix provides particular detail on five of the determinants that fall into this third group. This material in Appendix H represents both a compilation of existing literature as well as some new syntheses of recent studies; we commend the reader's attention to this important information.

OVERALL SUSCEPTIBILITY

Taken together, the evidence regarding the individual mediators of susceptibility described in Appendix H supports the plausibility of a continuous distribution of susceptibility in the human population. Some of the individual determinants of susceptibility, such as concentrations of activating enzymes or of proteins that might become oncogenic, may themselves exist in continuous gradations across the human population. Even factors that have long been thought to be dichotomous are now being revealed as more complicated—e.g., the recent finding that a substantial fraction of the population is heterozygous for ataxia-telangiectasia and has a susceptibility midway between that of ataxia-telangiectasia homozygotes and that of "normal" people (Swift et al., 1991). Most important, the combination of a large number of genetic, environmental, and lifestyle influences, even if each were bimodally distributed, would likely generate an essentially continuous overall susceptibility distribution. As Reif (1981) has noted, "we would expect to find in [the outbred human population] what would be the equivalent result of outbreeding different strains of inbred mice: a spectrum of different genetic predispositions for any particular type of tumor."

A working definition of the breadth of the distribution of "interindividual variability in overall susceptibility to carcinogenesis" is as follows: If we identified persons of high susceptibility (say, we knew them to represent the 99th percentile of the population distribution) and low susceptibility (say, the 1st percentile), we could estimate the risks that each would face if subjected to the same exposure to a carcinogen. If the estimated risk to the first type of person were 10^{-2} and the estimated risk to the second type of person were 10^{-6}, we could say that "human susceptibility to this chemical varies by at least a factor of 10,000."[4]

There are two distinct but complementary approaches to estimating the form and breadth of the distribution of interindividual variability in overall susceptibility to carcinogenesis. The biologic approach is a "bottom-up" method that uses empirical data on the distribution of particular factors that mediate susceptibility to model the overall distribution. In the major quantitative biologic analysis of the possible extent of human variations in susceptibility to carcinogenesis, Hattis et al. (1986) reviewed 61 studies that contained individual human data

[4]Similarly, the two persons might face equal cancer risks at exposures that were 10,000-fold different. However, an alternative definition, which would be more applicable for threshold effects, would be to call the difference in susceptibility the ratio of doses needed to produce the same effect in two different individuals.

on six characteristics that are probably involved causally in the carcinogenic process. The six were the half-life of particular biologically active substances in blood, metabolic activation of drugs (in vivo) and putative carcinogens (in vitro), enzymatic detoxification, DNA-adduct formation, the rate of DNA repair (as measured by the rate of unscheduled DNA synthesis induced by UV light), and the induction of sister-chromatid exchanges after exposure of lymphocytes to x-rays. They estimated the overall variability in each factor by fitting a lognormal distribution to the data and then propagated the variabilities by using Monte Carlo simulation and assuming that the factors interacted multiplicatively and were statistically independent. Their major conclusion was that the logarithmic standard deviation of the susceptibility distribution lies between 0.9 and 2.7 (90% confidence interval). That is, the difference in susceptibility between the most sensitive 1% of the population and the least sensitive 1% might be as small as a factor of 36 (if the logarithmic standard deviation was 0.9) or as large as a factor of 50,000 (if the logarithmic standard deviation was 2.7).[5]

The alternative approach is inferential or "top-down," and combines epidemiologic data with a demographic technique known as heterogeneity dynamics. Heterogeneity dynamics is an analytic method for describing the changing characteristics of a heterogeneous population as its members age. The power of the heterogeneity-dynamics approach to explain initially puzzling aspects of demographic data, as well as to challenge simplistic explanations of population behavior, stems from its emphasis on the divergence between forces that affect individuals and forces that affect populations (Vaupel and Yashin, 1983). The most fundamental concept of heterogeneity dynamics is that individuals change at rates different from those of the cohorts they belong to, because the passage of time affects the composition of the cohort as it affects the life prospects of each member. In a markedly heterogeneous population, the overall death rate can decline with age, even though every individual faces an ever-increasing risk of death, simply because the population as a whole grows increasingly more "resistant" to death as the more susceptible members are preferentially removed. Specifically with regard to cancer, heterogeneity dynamics can examine the progressive divergence of observed human age-incidence functions (for many tumor types) away from the function that is believed to apply to an individual's risk as a function of age—namely, the power function of age formalized in the 1950s by Armitage and Doll (which posits that risk increases proportionally with age raised to an integral exponent, probably 4, 5, or 6). In contrast with groups of inbred laboratory animals, which do exhibit age-incidence functions that generally obey the Armitage-Doll model, in humans the age-incidence curves for many tumor types begin to level off and plateau at higher ages.

Many of the pioneering studies that used heterogeneity dynamics to infer the amount of variation in human susceptibility to cancer used cross-sectional data, which might have been confounded by secular changes in exposures to carcinogenic stimuli (Sutherland and Bailar,

---

[5]The logarithmic standard deviation is equivalent to the standard deviation of the normal distribution corresponding to the particular lognormal distribution. If one takes the antilog of the logarithmic standard deviation, one obtains the "geometric standard deviation", or GSD, which has a more intuitively appealing definition: N standard deviations away from the median corresponds to multiplying or dividing the median by the GSD raised to the power N.

1984; Manton et al., 1986). One investigation that built on the previous body of work was that of Finkel (1987), who assembled longitudinal data on cancer mortality, including the age at death and cause of death of all males and females born in 1890, for both the United States and Norway. That study separately examined deaths due to lung cancer and colorectal cancer and tried to infer the amount of population heterogeneity that could have caused the observed age-mortality relationships to diverge from the Armitage-Doll (age^N) function that should apply to the population if all humans are of equal sensitivity. The study concluded that as a first approximation, the amount of variability (for either sex, either disease, and either country) could be roughly modeled by a lognormal distribution with a logarithmic standard deviation on the order of 2.0 (i.e., general agreement with the results of Hattis et al., 1986). That is, about 5% of the population might be about 25 times more susceptible than the average person (and a corresponding 5% about 25 times less susceptible); about 2.5% might be 50 times more (or less) susceptible than the average, and about 1% might be at least 100 times more (or less) susceptible.

A later analysis (Finkel, in press) showed that such a conclusion, if borne out, would have important implications not only for assessing risks to individuals, but for estimating population risk in practice. In a highly heterogeneous population, quantitative uncertainties about epidemiological inferences drawn from relatively small subpopulations (thousands or fewer), as well as the frequent application of animal-based risk estimates to similarly "small" subpopulations, will be increased by the possibility that the *average* susceptibility of small groups varies significantly from group to group.

The issue of susceptibility is an important one for acute toxicants as well as carcinogens. The NRC Committee on Evaluation of the Safety of Fishery Products addressed this issue in depth in their report entitled *Seafood Safety* (NRC, 1991b). Guidelines for the assessment of acute toxic effects in humans have recently been published by the NRC Committee on Toxicology (NRC, 1993d).

CONCLUSIONS

This section records the results of the committee's analysis of EPA's practice on variability.

EXPOSURE VARIABILITY AND THE MAXIMALLY EXPOSED INDIVIDUAL

One of the contentious defaults that has been used in past air-pollutant exposure and risk assessments has been the maximally exposed individual (MEI), who was assumed to be the person at greatest risk and whose risk was calculated by assuming that the person resided outdoors at the plant boundary, continuously for 70 years. This is a worst-case scenario (for exposure to the particular source only) and does not account for a number of obvious factors (e.g., the person spends time indoors, going to work, etc.) and other likely events (e.g., changing residence) that would decrease exposure to the emissions from the specific source.

This default also does not account for other, possibly countervailing factors involved in exposure variability discussed above. Suggestions to remedy this shortcoming have included decreasing the point estimate for residence time at the location to account for population mobility, and use of personal-activity models (see Chapters 3 and 6).

EPA's most recent exposure-assessment guidelines (EPA, 1992a) no longer use the MEI, instead coining the terms "high-end exposure estimates" (HEEE) and "theoretical upper-bounding exposure" (TUBE) (see Chapter 3). According to the new exposure guidelines (Section 5.3.5.1), a high-end risk "means risks above the 90th percentile of the population distribution, but not higher than the individual in the population who has the highest risk." The EPA Science Advisory Board had recommended that exposures or risks above the 99.9th percentile be regarded as "bounding estimates" (i.e., use of the 99.9th percentile as the HEEE) for large populations (assuming that unbounded distributions such as the lognormal are used as inputs for calculating the exposure or risk distribution). For smaller populations, the guidelines state that the choice of percentile should be based on the objective of the analysis. However, neither the HEEE nor the TUBE is explicitly related to the expected MEI.

The new exposure guidelines (Section 5.3.5.1) suggest four methods for arriving at an estimator of the HEEE. These are, in descending order of sophistication:

- "If sufficient data on the distribution of doses are available, take the value directly from the percentile(s) of interest within the high end;"
- "if. . . data on the parameters used to calculate the dose are available, a simulation (such as an exposure model or Monte Carlo simulation) can sometimes be made of the distribution. In this case, the assessor may take the estimate from the simulated distribution;"
- "if some information on the distribution of the variables making up the exposure or dose equation. . . is available, the assessor may estimate a value which falls into the high end. . . The assessor often constructs such an estimate by using maximum or near-maximum values for one or more of the most sensitive variables, leaving others at their mean values;"
- "if almost no data are available, [the assessor can] start with a bounding estimate and back off the limits used until the combination of parameter values is, in the judgment of the assessor, clearly in the distribution of exposure or dose. . . The availability of pertinent data will determine how easily and defensibly the high-end estimate can be developed by simply adjusting or backing off from the ultraconservative assumptions used in the bounding estimates."

The first two methods are much preferable to the last two and should be used whenever possible. Indeed, EPA should place a priority on collecting enough data (either case-specific or generic) that the latter two methods will not be needed in estimating variability in exposure. The distribution of exposures, developed from measurements or modeling results or both, should be used to estimate population exposure, as an input in calculating population risk. It can also be used to estimate the exposure of the maximally exposed person. For example, the most likely value of the exposure to the most exposed person is generally the 100[(N-1)/N]th percentile of the cumulative probability distribution characterizing interindividual variability in

exposures, where N is the number of persons used to construct the exposure distribution. This is a particularly convenient estimator to use because it is independent of the shape of the exposure distribution (see Appendix I-3). Other estimators of exposure to the highest, or jth highest for some $j<N$, person exposed are available (see Appendix I-3). The committee recommends that EPA explicitly and consistently use an estimator such as 100[(N-1)/N], because it, and not a vague estimate "somewhere above the 90th percentile," is responsive to the language in CAAA-90 calling for the calculation of risk to "the individual most exposed to emissions . . . ."

In recent times, EPA has begun incorporating into distributions of exposure assumptions that are based on a national average of years of residence in a home, as a replacement for its 70-year exposure assumption (e.g., an average lifetime). Proposals have been made for a similar "departure from default" for the time an individual spends at a residence each day, as a replacement for the 24 hours assumption. However, such analyses make the assumption that individuals move to a location of zero exposure when they change residences during their lifetime or leave the home each day. But, people moving from one place to another, whether it be changing the location of their residence or moving from the home to office, can vary greatly in their exposure to any one pollutant, from relatively high exposures to none. Furthermore, some exposures to different pollutants may be considered as interchangeable: moving from one place to another may yield exposures to different pollutants which, being interchangeable in their effects, can be taken as an aggregate, single "exposure." This assumption of interchangeability may or may not be realistic; however, because people moving from place to place can be seen as being exposed over time to a mixture of pollutants, some of them simultaneously and others at separate times, a simplistic analysis of residence times is not appropriate. The real problem is, in effect, a more complex problem of how to aggregate exposure to mixtures as well as one of multiple exposures of varying level of intensities to a single pollutant.

Thus, a simple distribution of residence times may not adequately account for the risks of movement from one region to another, especially for persons in hazardous occupations, such as agricultural workers exposed to pesticides, or persons of low socioeconomic status who change residences. Further, some subpopulations that might be more likely to reside in a high-exposure region might also be less mobile (e.g., owing to socioeconomic conditions). For these reasons, the default residency assumption for the calculation of the maximally exposed individual should remain at the mean of the current U.S. life expectancy, in the absence of supporting evidence otherwise. Such evidence could include population surveys of the affected area that demonstrate mobility outside regions of residence with similar exposures to similar pollutants. Personal activity (e.g., daily and seasonal activities) should be included.

If in a given case EPA determines that it must use the third method (combining various different "maximum," "near-maximum," and average values for inputs to the exposure equation) to arrive at the HEEE, the committee offers another caution: EPA has not demonstrated that these combinations of point estimates do in fact yield an output that reliably falls at the desired location within the overall distribution of exposure variability (that is, in the "conservative" portion of the distribution, but not above the confines of the entire distribu-

tion). Accordingly, EPA should validate (through generic simulation analyses and specific monitoring efforts) that its point-estimation methods do reasonably and reliably approximate what would be achieved via the more sophisticated direct-measurement or Monte Carlo methods (that is, a point estimate at approximately the 100[(N-1)/N]th percentile of the distribution). The fourth method, it should go without saying, is highly arbitrary and should not be used unless the bounding estimate can be shown to be "ultraconservative" and the concept of "backing off" is better defined by EPA.

SUSCEPTIBILITY

Human beings vary substantially in their inherent susceptibility to carcinogenesis, both in general and in response to any specific stimulus or biologic mechanism. No point estimate of the carcinogenic potency of a substance will apply to all individuals in the human population. Variability affects each step in the carcinogenesis process (e.g., carcinogen uptake and metabolism, DNA damage, DNA repair and misrepair, cell proliferation, tumor progression, and metastasis). Moreover, the variability arises from many independent risk factors, some inborn and some environmental. On the basis of substantial theory and some observational evidence, it appears that some of the individual determinants of susceptibility are distributed bimodally (or perhaps trimodally) in the human population; in such cases, a class of hypersusceptible people (e.g., those with germ-line mutations in tumor-suppressor genes) might be at tens, hundreds, or thousands of times greater risk than the rest of the population. Other determinants seem to be distributed more or less continuously and unimodally, with either narrow or broad variances (e.g., the kinetics or activities of enzymes that activate or detoxify particular pollutants).

To the extent that those issues have been considered at all with respect to carcinogenesis, EPA and the research community have thought almost exclusively in terms of the bimodal type of variation, with a normal majority and a hypersusceptible minority (ILSI, 1992). That model might be appropriate for noncarcinogenic effects (e.g., normal versus asthmatic response to SO_2), but it ignores a major class of variability vis-à-vis cancer (the continuous, "silent" variety), and it fails to capture even some bimodal cases in which hypersusceptibility might be the rule, rather than the exception (e.g., the poor-acetylator phenotype).

The magnitude and extent of human variability due to particular acquired or inherited cancer-susceptibility factors should be determined through molecular epidemiologic and other studies sponsored by EPA, the National Institutes of Health, and other federal agencies. Two priorities for such research should be

- To explore and elucidate the relationships between variability in each measurable factor (e.g., DNA adduct formation) and variability in susceptibility to carcinogenesis.
- To provide guidance on how to construct appropriate samples of the population for epidemiologic studies and risk extrapolation, given the influence of susceptibility variation on

uncertainty in population risk and the possible correlations between individual susceptibility and such factors as race, ethnicity, age, and sex.

Results of the research should be used to adjust and refine estimates of risks to individuals (identified, identifiable, or unidentifiable) and estimates of expected incidence in the general population.

The population distribution of interindividual variation in cancer susceptibility cannot now be estimated with much confidence. Preliminary studies of this question, both biologic (Hattis et al., 1986) and epidemiologic (Finkel, 1987) have concluded that the variation might be described as approximately lognormal, with about 10% of the population being different by a factor of 25-50 (either more or less susceptible) from the median individual (i.e., the logarithmic standard deviation of the distribution is approximately 2.0). While the estimated standard deviation of a susceptibility distribution suggested by these studies is uncertain, in light of the biochemical and epidemiological data reviewed earlier in this chapter it is currently not scientifically plausible that the U.S. population is strictly homogeneous in susceptibility to cancer induction by cancer-causing chemicals. EPA's guidelines are silent regarding person-to-person variations in susceptibility, thereby treating all humans as identical, despite substantial evidence and theory to the contrary. This is an important "missing default" in the guidelines. EPA does assume (although its language is not very clear in this regard) that the median human has susceptibility similar to that of the particular sex-strain combination of rodent that responds most sensitively of those tested in bioassays, or susceptibility identical with that of the particular persons observed in epidemiologic studies. These latter assumptions are reasonable as a starting point (Allen et al., 1988), but of course they could err substantially in either direction for a specific carcinogen or for carcinogens as a whole.

The missing default (variations in susceptibility among humans) and questionable default (average susceptibility of humans) are related in a straightforward manner. Any error of overestimation in rodent-to-human scaling (or in epidemiologic analysis) will tend to counteract the underestimation errors that must otherwise be introduced into some individual risk estimates by EPA's current practice of not distinguishing among different degrees of human susceptibility. Conversely, any error of underestimation in interspecies scaling will exacerbate the underestimation of individual risks for every person of above-average susceptibility. Therefore, EPA should increase its efforts to validate or improve the default assumption that the median human has similar susceptibility to that of the rodent strain used to compute potency, and should attempt to assess the plausible range of uncertainty surrounding the existing assumption. For further information, see the discussion in Chapter 11.

It can be argued, in addition, that EPA has a responsibility, insofar as it is practicable, to protect persons regardless of their individual susceptibility to carcinogenesis (we use *protect* here not in the absolute, zero-risk sense, but in the sense of ensuring that excess individual risk is within acceptable levels or below a *de minimus* level). It is unclear from the language in CAAA-90 Section 112(f)(2) whether the "individual most exposed to emissions" is intended to mean the person at highest risk when both exposure and susceptibility are taken into account, but this interpretation is both plausible and consistent with the fact that a major determinant of

susceptibility is the degree of metabolism of inhaled or ingested pollutants and the resulting exposure of somatic and germ cells to carcinogenic compounds (i.e., two people of different susceptibilities will likely be "exposed" to a different extent even if they breathe or ingest identical ambient concentrations). Moreover, EPA has a record of attempting to protect people with a combination of high exposure and high sensitivity, as seen in the National Ambient Air Quality Standards (NAAQS) program for criteria air pollutants (e.g., SO_2, NO_x, ozone, etc.).

Therefore, EPA should adopt an explicit default assumption for susceptibility before it begins to implement those decisions called for in the Clean Air Act Amendments of 1990 that require the calculation of risks to individuals. EPA could choose to incorporate into its cancer risk estimates for individual risk (not for population risk) a "default susceptibility factor" greater than the implicit factor of 1 that results from treating all humans as identical. EPA should explicitly choose a default factor greater than 1 if it interprets the statutory language to apply to individuals with both high exposure and above-average susceptibility.[6] EPA could explicitly choose a default factor of 1 for this purpose, if it interprets the statutory language to apply to the person who is average (in terms of susceptibility) but has high exposure. Or, preferably, EPA could develop a "default distribution" of susceptibility, and then generate the joint distribution of exposure and cancer potency (in light of susceptibility), to find the upper 95th or 99th percentile of risk for use in a risk assessment. The distribution is the more desirable way of dealing with this problem, because it takes explicit account of the joint probability (which may be large or small) of a highly exposed individual who is also highly susceptible.

Many of the currently known individual determinants of susceptibility vary by factors of hundreds or thousands at the cellular level; however, many of these risk factors (see Appendix I-2) tend to confer excess risks of approximately a factor of 10 on predisposed people, compared with "normal" ones. Although the total effect of the many such factors may cause susceptibility to vary upwards by more than a factor of 10, some members of the committee suggest that a default factor of 10 might be a reasonable starting point, if EPA wished to apply the statutory risk criteria (see Chapter 2) to the more susceptible members of the human

---

[6]Moreover, existing studies of overall variations in susceptibility suggest that a factor of 10 probably subsumes one or perhaps 1.5 standard deviations above the median for the normal human population. That is, assuming (as EPA does via its explicit default) that the median human and the rodent strain used to estimate potency are of similar susceptibility, an additional factor of 10 would equate the rodent response to approximately the 85th or 90th percentiles of human response. That would be a protective, but not a highly conservative, safety factor, inasmuch as perhaps 10 percent or more of the population would be (much) more susceptible than this new reference point.

Inclusion of a default factor of 10 could bring cancer risk assessment partway into line with the prevailing practice in noncancer risk assessment, wherein one of the factors of 10 that are often added is meant to account for person-to-person variations in sensitivity.

However, if EPA decides to use a factor of 10, it should emphasize that this is a default procedure that tries to account for some of the interindividual variation in dose-response relationships, but that in specific cases may be too high or too low to provide the optimum degree of "protection" (or to reduce risks to "acceptable" levels) for persons of truly unusual susceptibility. Nor does it ensure that (in combination with exposure estimates that might actually correspond to a maximally exposed or reasonably high-end person) risk estimates are predictive or conservative for the actual "maximally-at-risk" person. In contrast, some persons of extremely high susceptibility might, as a consequence of their susceptibility, not face high exposures. It might also be the case that some risk factors for carcinogenesis also predispose those affected to other diseases from which it might be impossible to protect them.

population. Conversely, other members of the committee do not consider an explicit factor of 10 to be justified at this time. A 10-fold adjustment might yield a reasonable best estimate of the high end of the susceptibility distribution for some pollutants when only a single predisposing factor divides the population into normal and hypersusceptible people.

If any susceptibility factor greater than 1 is applied, the short-term practical effect will be to increase all risk assessments for individual risk by the same factor, except for chemical-specific risk estimates where there is evidence that the variation in human susceptibility is larger or smaller for that chemical than for other substances. Such a general adjustment of either the default factor or default distribution might become appropriate when more information becomes available about the nature and extent of interindividual variations in susceptibility.

Individual risk assessments may depart from the new default when it can be shown either that humans are systematically either more or less sensitive than rodents to a particular chemical or that interindividual variation is markedly either more or less broad for this chemical than for the typical chemical. Therefore, in the spirit of our recommendations in Chapter 6 and Appendixes N-1 and N-2, the committee encourages EPA both to rethink the new default in general and to depart from it in specific cases when appropriately justified by general principles the agency should articulate.

Although it is known that there are susceptibility differences among people due to such factors as age, sex, race, and ethnicity, the nature and magnitude of these differences is not well known or understood; therefore, it is critical that additional research be pursued. As knowledge increases, science may be able to describe differences in the population at risk and recognize these differences with some type of default or distribution, although caution will be necessary to ensure that broad *correlations* between susceptibility and age, sex, etc., are not interpreted as deterministic *predictions*, valid for all individuals, or used in areas outside of risk assessment without proper respect for autonomy, privacy, and other social values.

In addition to adopting a default assumption for the effect of variations in susceptibility on individual risk, EPA should consider whether these variations might affect calculations of *population risk* as well. Estimates of population risk (i.e., the number of cases of disease or the number of deaths that might occur as a result of some exposure) are generally based on estimates of the average individual risk, which are then multiplied by the number of exposed persons to obtain a population risk estimate. The fact that individuals have unique susceptibilities should thus be irrelevant to calculating population risk, except if ignoring these variations *biases the estimate of average risk*. Some observers have pointed out a logical reason why EPA's current procedures might misestimate average risk. Even assuming that allometric or other interspecies scaling procedures correctly map the risk to test animals onto the "risk to the average human" (an assumption we encourage EPA to explore, validate, or refine), it is not clear *which* "average" is correctly estimated—the *median* (i.e., the risk to a person who has susceptibility at the 50th percentile of the population distribution) or the *expected value* (i.e., the average individual risk, taking into account *all* of the risks in the population and their frequency or likelihood of occurrence).

If person-to-person variation in susceptibility is small or symmetrically distributed (as in a

normal distribution), the median and the average (or mean) are likely to be equivalent, or so similar that this distinction is of no practical importance. However, if variation is large and asymmetrically distributed (as in a lognormal distribution with logarithmic standard deviation on the order of 2.0 or higher—see earlier example), the mean may exceed the median by roughly an order of magnitude or more.[7]

The committee encourages EPA to explore whether extrapolations made from animal bioassay data (or from epidemiological studies) at high exposures are likely to be appropriate for the median or for the average human, and to explore what response is warranted for the estimation and communication of population risk if the median and average are believed to differ significantly. As an initial position, EPA might assume that animal tests and epidemiological studies in fact lead to risk estimates for the *median* of the exposed group. This position would be based on the logic that at high exposures and hence high risks (that is, on the order of 10^{-2} for most epidemiologic studies, and 10^{-1} for bioassays), the effect of any variations in susceptibility within the test population would be truncated or attenuated. In such cases, any test animal or human subject whose *susceptibility* was X-fold higher than the median would face *risks* (far) less than X-fold higher than the median risk, because in no case can risk exceed 1.0 (certainty), and thus the effect of these individuals on the population average would not be in proportion to their susceptibilities. On the other hand, when extrapolating to ambient exposures where the median risk is closer to 10^{-6}, the full divergence between median and average in the general population would presumably manifest itself.

If, therefore, current procedures correctly estimate the median risk, then estimates of population risk would have to be increased by a factor corresponding to the ratio of the average to the median.

Other Changes in Risk-Assessment Methods

(1) Children are a readily identifiable subpopulation with its own physiologic characteristics (e.g., body weight), uptake characteristics (e.g., food consumption patterns), and inherent susceptibilities. When excess lifetime risk is the desired measure, EPA should compute an integrated lifetime risk, taking into account all relevant age-dependent variables, such as body weight, uptake, and average susceptibility (for one example of such a computation, see Appendix C of NRDC, 1989). If there is reason to believe that risk is not linearly related to biologically effective dose, and if the computed risks for children and adults are found to be significantly different, EPA should present separate risk assessments for children and adults.

(2) Although EPA has tried to take account of interindividual variability in susceptibility for non-cancer effects (e.g., in standards for criteria air pollutants such as ozone or SO_2), such efforts have neither seen exhaustive nor part of an overall focus on variability. In particular,

[7]For example, suppose the median income in a country was $10,000, but 5 percent of the population earned 25 times less or more than the median and an additional 1 percent earned 100 times less or more. Then the average income would be [(0.05)(400) + (0.05)(250,000) + (0.01)(100) + (0.01)(1,000,000) + (0.88)(10,000)] = $31,321, or more than three times the median income.

the "10-fold safety factor" used to account for interindividual variability when extrapolating from animal toxicity data has not been validated, in the sense that EPA is generally not aware how much of the human population falls within an order of magnitude of the median susceptibility for any particular toxic stimulus.

Although this chapter has focused on susceptibility to carcinogens, because this subject has received even less attention than that of susceptibility to non-carcinogens, the committee urges EPA to continue to improve its treatment of variability in the latter area as well.

(3) EPA has not sufficiently accounted for interindividual variability in biologic characteristics when it has used various physiologic or biologically based risk-assessment models. The validity of many of these models and assumptions depends crucially on the accuracy and precision of the human biological characteristics that drive them. In a wide variety of cases, interindividual variation can swamp the simple measurement uncertainty or the uncertainty in modeling that is inherent in deriving estimates for the "average" person. For example, physiologically based pharmacokinetic (PBPK) models require information about partition coefficients and enzyme concentrations and activities; Moolgavkar-Venzon-Knudson and other cell-kinetics models require information about cell growth and death rates and the timing of differentiation; and specific alternative models positing dose-response thresholds for given chemicals require information about ligand-receptor kinetics or other cellular phenomena. EPA has begun to collect data to support the development of distributions for the key PBPK parameters (such as alveolar ventilation rates, blood flows, partition coefficients, and Michaelis-Menten metabolic parameters) in both rodents and humans (EPA, 1988f). However, this database is still sparse, especially with respect to the possible variability in human parameters. EPA has developed point estimates for human PBPK parameters for 72 volatile organic chemicals, only 26 of which are on the list of 189 hazardous air pollutants covered in CAAA-90. For only five chemicals (benzene, *n*-hexane, toluene, trichloroethylene, and *n*-xylene) does EPA have any information on the presumed average and range of the parameters in the human population. It is perhaps noteworthy that in the one major instance in which EPA has revised a unit risk factor for a hazardous air pollutant on the basis of PBPK data (the case of methylene chloride), no information on the possible effect of human variability was used (EPA, 1987d; Portier and Kaplan, 1989).

Even when the alternative to the default model hinges on a qualitative, rather than a quantitative, distinction, such as the possible irrelevance to humans of the alpha-2μ-globulin mechanism involved in the initiation of some male rat kidney tumors, the new model must be checked against the possibility that some humans are qualitatively different from the norm. Any alternative assumption might be flawed, if it turns out to be biologically inappropriate for some fraction of the human population. Finally, although epidemiology is a powerful tool that can be used as a "reality check" on the validity of potency estimates derived from animal data, there must be a sufficient amount of human data for this purpose. The sample size needed for a study to have a given power level *increases* under the assumption that humans are not of identical susceptibility.

When EPA proposes to adopt an alternative risk-assessment assumption (such as use of a PBPK model, use of a cell-kinetics model, or the determination that a given animal response is

"not relevant to humans"), it should consider human interindividual variability in estimating the model parameters or verifying the assumption of "irrelevance." If the data are not available that would enable EPA to take account of human variability, EPA should be free to make any reasonable inferences about its extent and impact (rather than having to collect or await such data), but should encourage other interested parties to collect and provide the necessary data. In general, EPA should ensure that a similar level of variability analysis is applied to both the default and the alternative risk assessment, so that it can compare estimates of equal conservatism from each procedure.

Risk Communication

EPA often does not adequately communicate to its own decision-makers, to Congress, or to the public the variabilities that are and are not accounted for in any risk assessment and the implications for the conservatism and representativeness of the resulting risk numbers. Each of EPA's reports of a risk assessment should state its particular assumptions about human behavior and biology and what these do and do not account for. For example, a poor risk characterization for a hazardous air pollutant might say "The risk number R is a plausible upper bound." A better characterization would say, "The risk number R applies to a person of reasonably high-end behavior living at the fenceline 8 hours a day for 35 years." EPA should, whenever possible, go further and state, for example, "The person we are modeling is assumed to be of average susceptibility, but eats F grams per day of food grown in his backyard; the latter assumption is quite conservative, compared with the average."

Risk-communication and risk-management decisions are more difficult when, as is usually the case, there are both uncertainty and variability in key risk-assessment inputs. It is important, whenever possible, to separate the two phenomena conceptually, perhaps by presenting multiple analyses. For its full (as opposed to screening-level) risk assessments, EPA should acknowledge that all its risk numbers are made up of three components: the estimated risk itself (X), the level of confidence (Y) that the risk is no higher than X, and the percent of the population (Z) that X is intended to apply to in a variable population. EPA should use its present practice of saying that "the plausible upper-bound risk is X" only when it believes that Y and Z are both close to 100%. Otherwise, it should use statements like, "We are Y% certain that the risk is no more than X to Z% of the population," or use an equivalent pictorial representation (See Figure 10-2).

As an alternative or supplement to estimating the value of Z, EPA can and should try to present multiple scenarios to explain variability. For example, EPA could present one risk number (or preferably, an uncertainty distribution—see Chapter 9) that explicitly applies to a "person selected at random from the population," one that applies to a person of reasonably high susceptibility but "average" behavior (mobility, breathing rate, food consumption, etc.), and one that applies to a person whose susceptibility and behavioral variables are both in the "reasonably high" portion of their distributions.

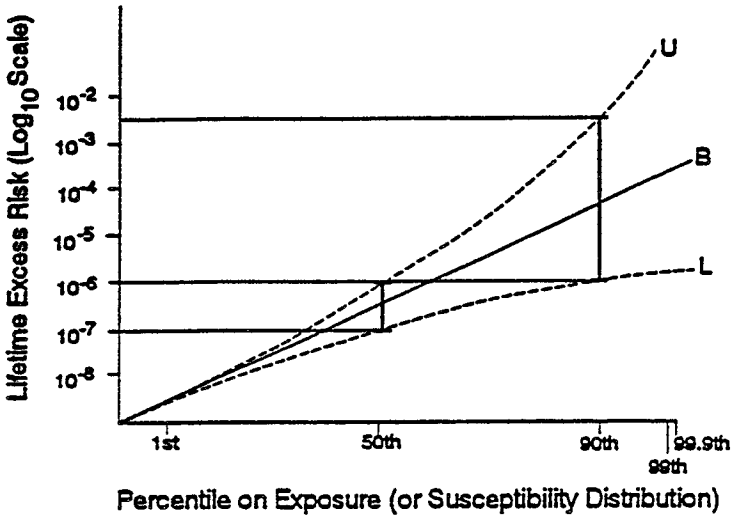


Figure 10-2. Communicating Risk, Uncertainty, and Variability Graphically

- *Curve B* presents the best estimate of the relationship between exposure (or susceptibility) and risk [expressed in a scale relative to the rest of the population, not in absolute exposure (or susceptibility) units.]

- *Curve L* presents the 5th (or other lower) percentile of this relationship.

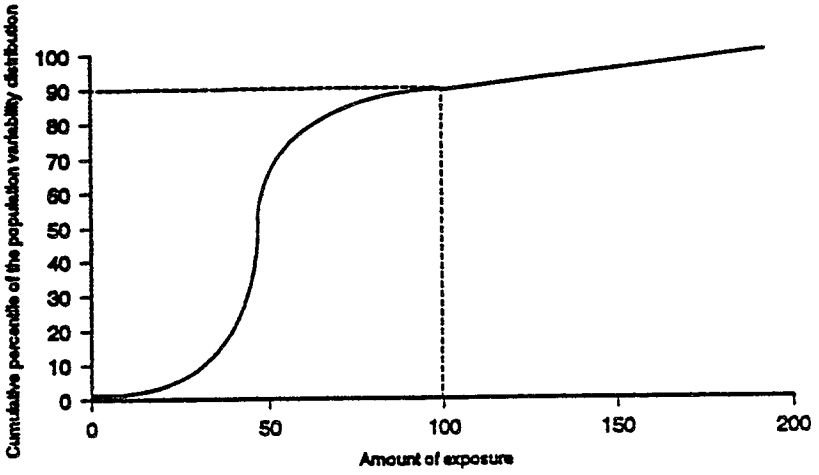
- *Curve U* presents the 95th (or other upper) percentile of this relationship.

Thus, for this hypothetical example, the risk communicator could say:

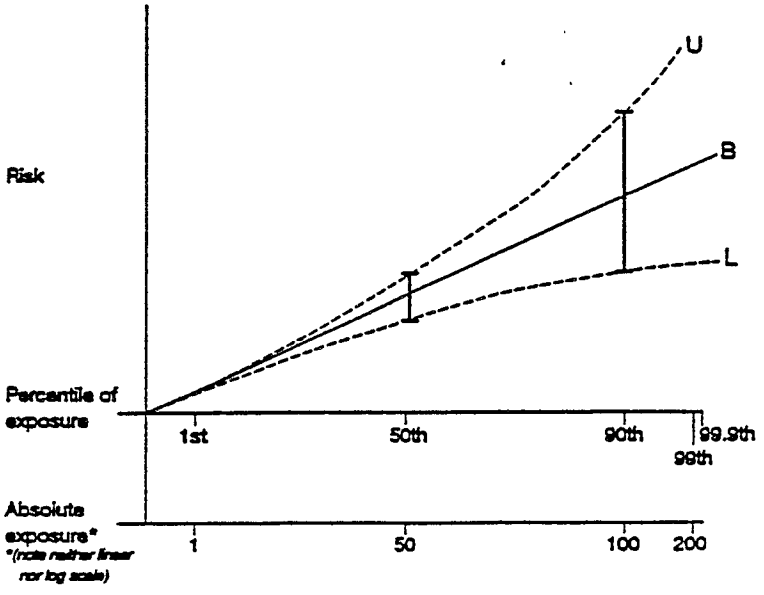
"We are 90% certain that the risk to the person with median exposure is between 10^{-7} and 10^{-6},"
AND/OR

"We are 90% certain that the risk to the person with high (90th percentile) exposure is between 10^{-6} and 3×10^{-3}."

Note: To translate from percentile-relative exposure to absolute exposure, you could add a second x-axis scale based on a figure as follows:



Thus, combining figures 1 and 2 gives you the 2 x-axes:



Not all the suggestions presented here, especially those regarding variation in susceptibility, might apply in every regulatory situation. The committee notes that in the past, whenever persons of high risk or susceptibility have been identified, society has tended to feel a far greater responsibility to inform and protect them. For such identifiable variability, the recommendations in this section are particularly salient. However, interindividual variability might be important even when the specific people with high and low values of the relevant characteristic cannot currently be identified.[8] Regardless of whether the variability is now identifiable (e.g., consumption rates of a given foodstuff), difficult to identify (e.g., presence of a mutant allele of a tumor-suppressor gene), or unidentifiable (e.g., a person's net susceptibility to carcinogenesis), the committee agrees that it is important to think about its potential magnitude and extent, to make it possible to assess whether existing procedures to estimate average risks and population incidence are biased or needlessly imprecise.

In contrast with issues involving average risk and incidence, however, some members of the committee consider the distribution of individual susceptibilities and the uncertainty as to where each person falls in that distribution to be irrelevant if the variation is and will remain unidentifiable. For example, some argue that people should be indifferent between a situation wherein their risk is determined to be precisely 10^{-5} or one wherein they have a 1% chance of being highly susceptible (with risk = 10^{-3}) and a 99% chance of being immune, with no way to know which applies to whom. In both cases, the expected value of individual risk is 10^{-5}, and it can be argued that the distribution of risks is the same, in that without the prospect of identifiability no one actually faces a risk of 10^{-3}, but just an equal chance of facing such a risk (Nichols and Zeckhauser, 1986).

Some of the members also argue that as we learn more about individual susceptibility, we will eventually reach a point where we will know that some individuals are at extremely high risk (i.e., carried to its extreme, an average individual risk of 10^{-6} may really represent cases where one person in each million is guaranteed to develop cancer while everyone else is immune). As we approach this point, they contend, society will have to face up to the fact that in order to guarantee that everyone in the population faces "acceptable" low levels of risk, we would have to reduce emissions to an impossibly low extent.

Other committee members reject or deem irrelevant the notion that risk is ultimately either zero or 1; they believe that, both for an individual's assessment of how foreboding or tolerable a risky situation is and for society's assessment of how just or unjust the distribution of risks is, the information about the unidentifiable variability must be reported—that it affects both judgments. To bolster their contentions, these members cite literature about the limitations of expected utility theory, which takes the view, contradicted by actual survey data, that the distribution of risky outcomes about their mean values should not affect the individual's

[8]"Currently" is an important qualifier given the rapid increases in our understanding of the molecular mechanisms of carcinogenesis. During the next several decades, science will doubtless become more adept at identifying individuals with greater susceptibility than average, and perhaps even pinpoint specific substances to which such individuals are particularly susceptible.

evaluation of the situation (Schrader-Frechette, 1985; Machina, 1990), and empirical findings that the skewness of lotteries over risky outcomes matters to people even when the mean and variance are kept constant (Lopes, 1984). They also argue that EPA should maintain consistency in how it handles exposure variability, which it reports even when the precise persons at each exposure level cannot be identified; i.e., EPA reports the variation in air concentration and the maximal concentration from a source even when (as is usually the case) it cannot predict exactly where the maximum will occur. If susceptibility is in large part related to person-to-person differences in the amount of carcinogenic material that a person's cells are exposed to via metabolism, then it is essentially another form of exposure variability, and the parallel with ambient (outside-the-body) exposure is close. Finally, they claim that having agreed that issues of pure uncertainty are important, EPA (and the committee) must be consistent and regard unidentifiable variability as relevant (see Appendix I-3). Our recommendations in Chapter 9 reflect our view that uncertainty is important because individuals and decision-makers do regard values other than the mean as highly relevant. If susceptibility is unidentifiable, then to the individual it represents a source of uncertainty about his or her individual risk, and many members of the committee believe it must be communicated just as uncertainty should be.

Social-science research aimed at clarifying the extent to which people care about unidentifiable variability in risk, the costs of accounting for it in risk management, and the extent to which people want government to take such variation and costs into account in making regulatory decisions and in setting priorities might be helpful in resolving these issues.

FINDINGS AND RECOMMENDATIONS

The committees findings and recommendations are briefly summarized below.

EXPOSURE

Historically, EPA has defined the maximally exposed individual (MEI) as the worst-case scenario—a continuous 70-year exposure to the maximal estimated long-term average concentration of a hazardous air pollutant. Departing from this practice, EPA has recently published methods for calculating bounding and "reasonably high-end" estimates of the highest actual or possible exposures using a real or default distribution of exposure within a population. The new exposure guidelines do not explicitly define a point on this distribution corresponding to the highest expected exposure level of an individual.

- The committee endorses the EPA's use of bounding estimates, but only in screening assessments to determine whether further levels of analysis are necessary. For further levels of analysis, the committee supports EPA's development of distributions of exposure values based on available measurements, modeling results, or both. These distributions can also be

used to estimate the exposure of the maximally exposed person. For example, the most likely value of the exposure to the most exposed person is generally the 100[(N - 1)/N]th percentile of the cumulative probability distribution characterizing interindividual variability in exposure, where N is the number of persons used to construct the exposure distribution. This is a particularly convenient estimator to use because it is independent of the shape of the exposure distribution. The committee recommends that EPA explicitly and consistently use an estimator such as 100[(N - 1)/N], because it, and not a vague estimate "somewhere above the 90th percentile," is responsive to the language in CAAA-90 calling for the calculation of risk to "the individual most exposed to emissions . . . ."

In recent times, EPA has begun incorporating into distributions of exposure assumptions that are based on a national average of years of residence in a home, as a replacement for its 70-year exposure assumption (e.g., an average lifetime). Proposals have been made for a similar "departure from defaults" for the time an individual spends at a residence each day, as a replacement for the 24 hours assumption. However, such analyses make the assumption that individuals move to a location of zero exposure when they change residences during their lifetime or leave the home each day. But, people moving from one place to another, whether it be changing the location of their residence or moving from the home to office, may vary greatly in their exposure to any one pollutant, from relatively high exposures to none. Further, some exposures to different pollutants may be considered as interchangeable: moving from one place to another may yield exposures to different pollutants which, being interchangeable in their effects, can be taken as an aggregate, single "exposure." This assumption of interchangeability may or may not be realistic; however, because people moving from place to place can be seen as being exposed, over time to a mixture of pollutants, some of them simultaneously and others at separate times, a simplistic analysis of residence times is not appropriate. The real problem is, in effect, a more complex problem of how to aggregate exposure to mixtures as well as one of multiple exposures of varying level of intensities to a single pollutant. Thus, a simplistic analysis based on a simple distribution of residence times is not appropriate.

- EPA should use the mean of current life expectancy as the assumption for the duration of individual residence time in a high-exposure area, or a distribution of residence times which accounts for the likelihood that changing residences might not result in significantly lower exposure. Similarly, EPA should use a conservative estimate for the number of hours a day an individual is exposed, or develop a distribution of the number of hours per day an individual spends in different exposure situations. Such information can be gathered through neighborhood surveys, etc. in these high-exposure areas. Note that the distribution would correctly be used only for individual risk calculations, as total population risk is unaffected by the number of persons whose exposures sum to a given total value (if risk is linearly related to exposure rate).

EPA has not provided sufficient documentation in its exposure-assessment guidelines to

ensure that its point-estimation techniques used to determine the "high-end exposure estimate" (HEEE) when data are sparse reliably yield an estimate at the desired location within the overall distribution of exposure (which, according to these guidelines, lies above the 90th percentile but not beyond the confines of the entire distribution).

- EPA should provide a clear method and rationale for determining *when* point estimators for the HEEE can or should be used instead of a full Monte Carlo (or similar) approach to choosing the desired percentile explicitly. The rationale should more clearly indicate how such estimators are to be generated, should offer more documentation that such point-estimation methods do yield reasonably consistent representations of the desired percentile, and should justify the choice of such a percentile if it differs from that which corresponds to the expected value of exposure to the "person most exposed to emissions.".

POTENCY

EPA has dealt little with the issue of human variability in susceptibility; the limited efforts to date have focused exclusively on variability relative to noncarcinogenic effects (e.g., normal versus asthmatic response to SO_2). The appropriate response to variability for noncancer end points (i.e., identify the characteristics of "normal" and "hypersusceptible" individuals, and then decide whether or not to protect both groups) might not be appropriate for carcinogenesis, in which variability might well be continuous and unimodal, rather than either-or.

- EPA, NIH, and other federal agencies should sponsor molecular epidemiologic and other research on the extent of interindividual variability in various factors that affect susceptibility and cancer, on the relationships between variability in each factor and in the health end point, and on the possible correlations between susceptibility and such covariates as age, race, ethnicity, and sex. Results of the research should be used to adjust and refine estimates of risks to individuals (identified, identifiable, or unidentifiable) and estimates of expected incidence in the general population. As this research progresses, the natural science and social science community should collaborate to explore the implications of any susceptibility factors that can be tested for or that strongly correlate with other genetic traits, so as to ensure that any findings are not misinterpreted or used outside of the environmental risk assessment arena without proper care.

SUSCEPTIBILITY

EPA does not account for person-to-person variations in susceptibility to cancer; it thereby treats all humans as identical in this respect in its risk calculations.

- EPA should adopt a default assumption for susceptibility before it begins to implement

those decisions called for in the Clean Air Act that require the calculation of risks to individuals. EPA could choose to incorporate into its cancer risk estimates for individual risk a "default susceptibility factor" greater than the implicit factor of 1 that results from treating all humans as identical. EPA should explicitly choose a default factor greater than 1 if it interprets the statutory language to apply to an individual with high exposure and above-average susceptibility. EPA could explicitly choose a default factor of 1 for this purpose, if it interprets the statutory language to apply to an individual with high exposure but average susceptibility. Preferably, EPA could develop a "default distribution" of susceptibility, and then generate the joint distribution of exposure and cancer potency (in light of susceptibility) to find the upper 95th percentile (or 99th percentile) of risk for each risk assessment.

EPA makes its potency calculations on the assumption that, on average, humans have susceptibility similar to that of the particular sex-strain combination of rodent that responds most sensitively of those tested in bioassays or susceptibility identical with that of the particular groups of persons observed in epidemiologic studies.

- EPA should continue and increase its efforts to validate or improve the default assumption that, on average, humans to be protected at the risk-management stage have susceptibility similar to that of humans included in relevant epidemiological studies, the most-sensitive rodents tested, or both.

It is possible that ignoring variations in human susceptibility may cause significant underestimation of population risk, if both of two conditions hold: (1) current procedures to extrapolate results of laboratory bioassays or epidemiologic studies to the general population correctly map the observed risk in the test population to the human with median susceptibility, not to the expected value averaged over the entire general population; and (2) there is sufficient skewed variability in susceptibility in the general population to cause the expected value to exceed the median to a significant extent.

- In addition to continuing to explore the assumption that interspecies scaling (or epidemiologic extrapolation) correctly predicts average human susceptibility, EPA should investigate whether the average that is predicted corresponds to the median or the expected value. If there is reason to suspect the former is true, EPA should consider whether it needs to adjust its estimates of population risk to account for this discrepancy.

Children are a readily identifiable subpopulation with its own physiologic characteristics (e.g., body weight), uptake characteristics (e.g., food consumption patterns), and inherent susceptibilities.

- If there is reason to believe that risk of adverse biological effects per unit dose depends on age, EPA should present separate risk estimates for adults and children. When excess lifetime

risk is the desired measure, EPA should compute an integrated lifetime risk, taking into account all relevant age-dependent variables.

EPA does not usually explore or consider interindividual variability in key biologic parameters when it uses or evaluates various physiologic or biologically based risk-assessment models (or else evaluates some data but does not report on this in its final public documents). In some other cases, EPA does gather or review data that bear on human variability, but tends to accept them at face value without ensuring that they are representative of the entire population. As a general rule, the larger the number of characteristics with an important effect on risk or the more variable those characteristics are, the larger the sample of the human population needed to establish confidently the mean and range of each of those characteristics.

- When EPA proposes to adopt an alternative risk-assessment assumption (such as use of a PBPK model, use of a cell-kinetics model, or the determination that a given animal response is "not relevant to humans"), it should consider human interindividual variability in estimating the model parameters or verifying the assumption of "irrelevance." If the data are not available to take account of human variability, EPA should be free to make any reasonable inferences about its extent and impact (rather than having to collect or await such data), but should encourage other interested parties to collect and provide the necessary data. In general, in parallel to recommendation UAR4, EPA should ensure that a similar level of variability analysis is applied to both the default and the alternative risk assessment, so that it can compare equivalently conservative estimates from each procedure.

RISK COMMUNICATION

EPA does not adequately communicate to its own decision-makers, to Congress, or to the public the variabilities that are and are not accounted for in any risk assessment and the implications for the conservatism and representativeness of the resulting risk numbers.

- EPA should carefully state in each risk assessment what its particular assumptions about human behavior and biology do and do not account for.

For its full (as opposed to screening-level) risk assessments, EPA makes risk-communication and risk-management decisions more difficult when, as is usually the case, both uncertainty and variability are important.

- Whenever possible, EPA should separate uncertainty and variability conceptually, perhaps by presenting multiple analyses. EPA should acknowledge that all its risk numbers are made up of three components: the estimated risk itself (X), the level of confidence (Y) that the risk is no higher than X, and the percent of the population (Z) that X is intended to apply to in a

variable population. In addition, rather than reporting both Y and Z, EPA can and should try to present multiple scenarios to explore and explain the variability dimension.